International Journal on
"Technical and Physical Problems of Engineering"
(IJTPE)
Published by International Organization of IOTPE

ISSN 2077-3528
IJTPE Journal
www.iotpe.com
ijtpe@iotpe.com

June 2024 | Issue 59 | Volume 16 | Number 2 | Pages 383-389

FAKE NEWS DETECTION: A DATA MINING PERSPECTIVE

M. El Mohadab [1] B. Bouikhalene [2] S. Safi [3]

1. LAROSERI, Computer Science Department, Faculty of Sciences, University of Chouaib Doukkali, El Jadida, Morocco, m.elmohadab@gmail.com
2. Mathematics and Computer Science Department, Polydisciplinary Faculty, University of Sultan Moulay Slimane, Beni Mellal, Morocco, b.bouikhalene@usms.ma, safisaid@gmail.com

Abstract- News encompasses information pertaining to current events and is disseminated through various medias such as word of mouth, print, postal services, broadcasting, electronic communication, and firsthand accounts of witnesses or observers. It covers a broad range of topics including education, health, environment, economy, fashion and sports. Now days, the information is shared in the internet from different source, where the emergence of fake news which has spread widely on the internet because so much information is shared about the media. Social media platforms, with their easy access to information and its rapid dissemination encourage people to seek and consume it, on social media because they represent an important source to disseminate information of all kinds the latter use preventive measures to fight against using reporting and detection tools. There are currently a variety of methods for identifying fake news utilizing machine learning. In this paper, we resort to making use of artificial intelligence through machine learning; which can motivated our study to determine the best performing algorithm between one of the algorithms for machine learning: Decision Tree, Support vector machine, Logistic Regression and K-nearest neighbor by working on the basis of Fake news data from the kaggle website. The Support vector machine algorithm showed good results, they reached more than 92% accuracy.

Keywords: Classification Algorithms, Data Mining, Fake News, Fake News Detection, Machine Learning.

1. INTRODUCTION

For some years now, the emergence and spread of fake news [1] in many sectors has been a major concern for many institutions and companies, so research into the detection of misleading news is trying to address the as associated issue at an early phase, as it could be a recent development, related to the public concern and interest experienced recently in the emergence of the covid-19 pandemic [2]. In general, such fake news is invented by the authors either in a context to disseminate totally erroneous information or to amuse the reader, or through poor writing or miscommunication of real news these will be poorly acquired by the target audience, research into the detection of fake news is a new area of study.

Because of these factors, there is no universally accepted definition of fake news, has not yet been created. There are several methods discussed in the literature, each with its own advantages and disadvantages, the problem of fake news detection is analyzed in depth. However, in such a way, from the study of the literature, the issue of false information detection tries to achieve higher performance. Changing the level of confidence at the database, historical data, and profile levels can be improve. By applying the trust indicator at different stage, the performance of false information detection and access control can be enhanced. This document is split into eight sections. Section 2 gives an overview of the detection of fake news. Section 3 introduces the basics of learning methods, which are required for understanding how to choose the best classifier in learning method. The method proposed is shown in Section 4. In this section, the dataset description and the dataset preprocessing with detailed steps is explained. Section 5 presents the basics of the algorithms used in this study Section 6, results in terms of different metrics for each methods are presented. In conclusion, a few points will be given in Section 7.

2. RELATED WORKS

Because of the exponential growth of news spread by social media. Some of the news disseminated are fake, these have a negative impact either on society, finance, politics, etc.

To detect these fake news, numerous studies have been conducted employing methods of machine learning, among which we find:

Reis, et al. [3] added new features for training and training super-vised learning classifiers, However, the authors utilized a small data set method.

Goldani, et al. [4] proposed the application of neural network, distinct method of feature extraction, non-static integration during the learning phase. However, it is limited to only using textual features to classify or not classify fake news.

Liu and Wu [5] proposed an application of neural networks CNN by training on both positive and unlabeled samples, using five-way cross validation. However, the study is limited by a small dataset which restricts the scope of the proposed work.

Registered: 231205 | Article Info: 1869 | Accepted: 240517
Reserved: 240304 | Available: 240630

Vicario, et al. [6] proposed a new method to predict false news through social media using a classifier that is based on different features including semantic, structural, sentiment, predicted and user-based features, but this technique does not identify the element with negative impact on the news.

Asghar, et al. [7] used a hybrid neural network to specify if twittering are false or not and thus by approving them with distinct characteristics but the last one is restricted to evaluating fake news based solely on text.

Vereshchaka, et al. [8] used models from deep learning to classify binary data, such as LSTM, RNN, GRU, but it doesn't achieve greater accuracy, confined to a small data set.

Faustini and Covoes [9] used diverse evaluations method to detect fake news in each dataset by using custom features, DCDistance, Word2Vec, bag-of-words and in order to categorize, but the technique does not achieve high accuracy in terms of evaluation.

Rastogi and Bansal [10] utilized a geometry model of deep learning which combines between convolutional layers and two-dimensional features for their proposed work, which will be improved by adding topic-based news classification and early detection of fake news. Kaliyar, et al. [11] used the matrix-tenor to find features. Also, extracted characteristics are used for classification purposes DNN and XGBoost, however, it doesn't show how content and context are used in feature extraction.

Jadhav and Thepade [12] used the DSSM-LSTM semantic elements to detect fake news by categorizing information, but a failure to display feature extraction using semantic features is demonstrated.

Fake news has become one of the biggest problems of our time. Fake news represents a clear and present danger for democracy and the stability of society.

Social networks are a great space for content production, and between all the content where users can access, news is a most frequently accessed element. The information may be published by legislators, news channels, sites or even by common individuals. The originality of these messages must be verified, as the spread of false information is a real concern nowadays, and many companies are taking steps to educate the public about the consequences of spreading false information. Because manually categorizing information is a laborious process, it is impossible to accurately measure the originality of information available online.

Faced with the mass of information available (on social networks, on the internet, in newspapers, on television...) it multiplies drastically and it becomes complicated to make the difference between real and fake information.

In this study, we can train classification algorithms to classify news if they are true or false.

3. MACHINE LEARNING ALGORITHMS

Section 3 is reserved present an overview of the re-search regarding the models we worked with. There are four different types of machine learning [13]: supervised learning, which needs class data to be labelled before the model can use it to train; The second approach is unsupervised learning (clustering), which requires unlabeled input of information; the third type is semi-supervised learning where data can be manipulated with and without labels. And the last is reinforcement learning, which consists of letting an algorithm learns from its mistakes to reach a given goal. The machine learning types is described in Figure 1.

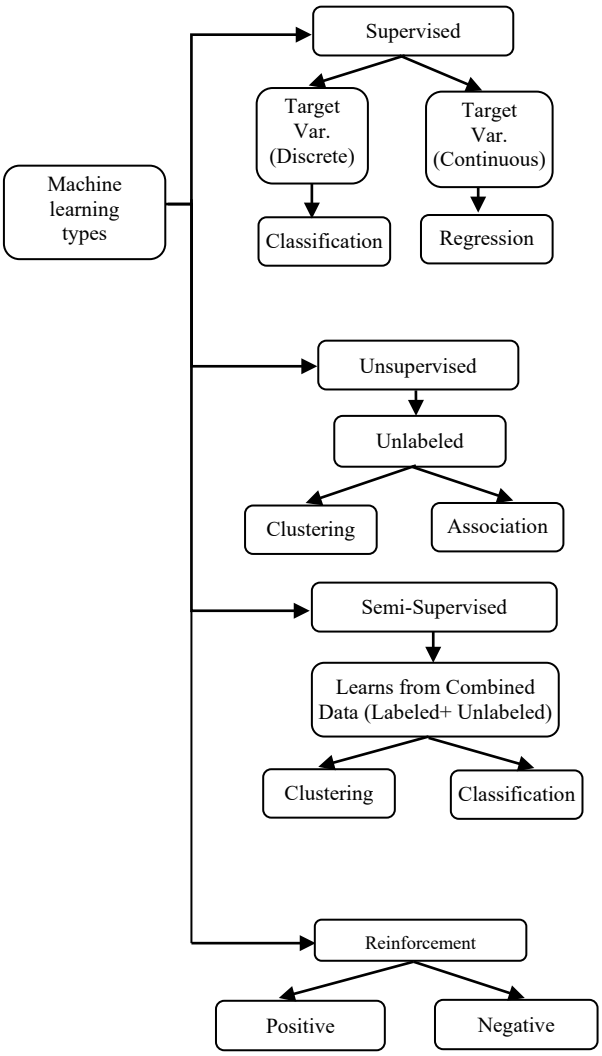

Figure 1. Different types of machine learning

3.1. Supervised Learning Method

Different supervised learning generally begins with a defined set of data and some comprehension of the classification system for the data. Supervised learning aims at finding models in data which can be applied to an analytical process [14].

3.2. Supervised Learning Method

The data is divided up into clusters or feature groups by unsupervised learning algorithms. The parameter values and data classification are created by the unlabeled data. This procedure basically gives the data labels to make it supervised. When there is a large amount of data, unsupervised learning can determine the outcome [15].

3.3. Semi Supervised Learning Method

Is a mixture of supervised and unsupervised techniques which works with labelled and unlabeled data in the same way. Thus, it falls between unsupervised learning and supervised learning. In real-world scenarios, labeled data can often be scarce across various contexts, while unlabeled data is abundant. In such cases, semi-supervised learning becomes invaluable [16].

3.4. Reinforcement Learning Method

Is a kind of algorithm used in machine learning. It consists to generate actions in response to the environment and finds errors or rewards. Trial, error and delayed rewards are the most significant aspects of reinforcement learning. [17].

4. PROPOSED METHODOLOGY

The main idea of our study is to build a model which can accurately predict if an article is true or false.

In the proposed study, the following is a brief description of the steps involved:

• Datasets Fake News were found and investigated for the proposed method.
• Preprocessing was done in order to get rid of the noise in the datasets.
• Tokenization is carried out in order to translate the larger text into words.
• Extracting the features for detecting fake news rather than text.
• To prepare the model, extracted features are passed.
• The efficiency of proposed models can be measured using evaluation metrics.

4.1. Dataset Description

The main objective of the study is to make a distinction between real news and fake news, have explored the database that contains a total of around 40,000 records which include both true and false news, the latter is downloaded from Kaggle [18]. We aim to improve the model by training it to accurately predict whether a news item is true or false.

Table1 presents a sample of the database used.

Table 1. Extract from the "Fake news" database

Id	Title	Text	Category
0	As U.S.budget fights looms, Republicans fli…	Washington (Reuters) - The head of a conservation	1
1	U.S. military to accept transgender recruits...	Washington (Reuters) - Transgender people ...	1
2	FBI Russia probe helped by Australian diplo...	Washington(Reuters) - Trump campaign advis …	1
3	McPain: John McCain Furious That Iran Tre...	21st Century wire says As 21WIRE reported ...	0
4	How to Blow $700 Million: Al Jazeera Ame...	21st Century wire says Al Jazeera America will ...	1
5	Sunistan: US and Allied 'Safe Zone' Plan to ...	Patrick Henningsen 21st Century wire Remember ...	0
6	JUSTICE? Yahoo Settles E-mail Privacy Cla..	21st Century wire says it is a familiar theme ...	0
7	10 U.S. Navy Sailors Held By Iranian Milit...	21st Century wire says As 21WIRE predicted in ...	0

The database retrieved from contains four columns each one of them has an information:

• Id: Which indicate the identifier of the news.
• Title: is the main title within the news is published.
• Text : the details about the news.
• Category field of the news: which indicate if the new is true or fales, it is a binary value 0 for fake news and 1 for true ones.

The fake and real news data are provided in two separate datasets, each dataset containing approximately 20,000 articles. We used binary labelling, with 0 and 1 corresponding respective to fake and real news.

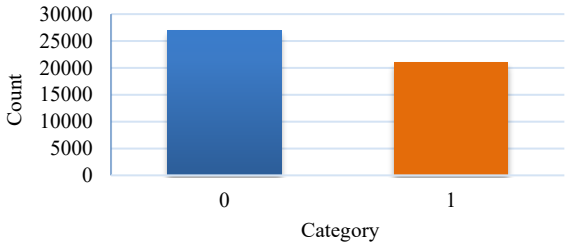

Figure 2. Fake and true data category

The four feature selection outputs in the form of train and test input are provided as input to the four machine learning classification algorithms, which produce results. The classification algorithm used are defined in the next section. The dataset retrieved is divided in two: 70% for preparation and 30% for testing. As training and testing input, the vocabulary matrix or bag of words that resulted is provided. The four machine learning algorithm calculation is then used to produce separate outcomes, which are then analyzed. Figure 3 presents the architecture diagram for detecting fake news.

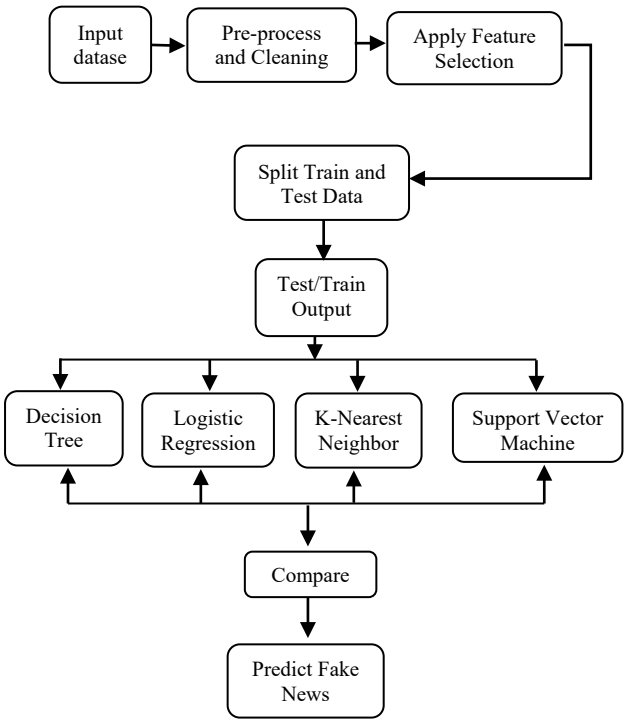

Figure 3. Architecture diagram for detecting fake news

4.2. Data Preprocessing

Virtually all types of data analysis, data science, or AI development require preliminary data preprocessing to provide reliable, accurate, and robust results. Applying data mining algorithms on raw data would not yield quality results because they would fail to effectively identify patterns, so data processing is important to improve overall data quality.

Using text data directly is difficult because it contains unusable words, special symbols and other things. If we use them directly without cleaning them, it will be very difficult for the algorithm to detect patterns in the text and sometimes it will also generate an error. So cleaning the data from the text is a critical step; The data needed to train machine learning or deep learning algorithms need to be preprocessed automatically.

The reorganization of raw data into an appropriate format for algorithms is assisted by feature engineering practices like data management, transformation, reduction, selection, and scaling. The pre-processing stage involved removing noise from datasets, including punctuation marks and HTML tags.

The following is a brief description of the steps involved in the proposed study:

• Delete lines with missing values.
• Convert all text to lower case.
• Delete empty words.
• Take only the words from the text and remove numbers and special characters.
• Tokenization: Convert the word or token to its basic form.

5. ALGORITHMS

The data is tested using various machine learning algorithms, which which represented as follow:

1- Decision Tree (DT) [19]: is a predictive modellint tool which can be utilized to solve regression and classification issues, but its primary purpose is to do so. A tree classifier is a predictive modeling tool utilizes internal nodes to represent dataset features, branches to denote decision rules, and leaf nodes to indicate the resulting outcomes. The leaf node and the decision node are both parts of a decision tree. Leaf nodes are the decisions' output and lack branches, whereas decision nodes are used to take a single decision and have multiple branches. Which is described by:

• Determine what type of node is terminal or whether it needs to be labeled as a leaf.
• Choose a test to coordinate with a node.
• Give each sheet a class.

2- Logistic Regression (LR) [20]: is a machine learning algorithm employs the supervised learning approach. Using a particular set of independent variables, it is used to predict the categorical dependent variable. It predicts how a categorical dependent variable will behave. The result should subsequently be a discrete or categorical value. It might be false or true, no or yes, zero or one, and so on. Instead of providing specific values like 0, 1, it explains probabilistic qualities that are somewhere in the range of 0, 1, which are described by:

• In ordinal logistic regression, the dependent variable must be ordinal, whereas in binary logistic regression, it must be binary.
• Repeated measurements or matched data should not be used as sources of observation.
• Logistic regression rejects multicollinearity as a separate variable.
• Operates under the assumption of independent linearity. It expects that independent variables are linearly connected to the log odds.

3- K-Nearest Neighbor (K-NN) [21] is an algorithm able of addressing both classification and regression tasks." K-Nearest Neighbors is a fundamental algorithm uses a similarity measure to predict the classification of unlabeled data and stores all of the available data. When plotting two parameters on the 2D Cartesian system in linear geometry, we determine the similarity measure by calculating the distance between the points. A similar applies here, the KNN algorithm deals with the suspicion that comparative things exist in close proximity, basically we can place in exactly the same things stay near one another. Which is described by:

• Decide on K as the number of neighbors.
• Compute the Euclidean or Manhattan distance from the unclassified point to the other points.
• Get the nearest neighbor K based on the estimated distance.
• Include the quantity of focuses in every category among these K neighbors.
• The most recent category among these K neighbors should receive the new point.
• Our model is ready.

4- Support Vector Machine (SVM) [22]: is an algorithm used for regression and classification problems. It is mostly used for classification problems in machine learning. The best boundary or decision that can divide the n-dimensional space into classes is the objective of the SVM algorithm. It will make it simple to put a new data point into the good category. But the limit is known as a hyperplane. Which is described by:

• Linear, polynomial, or Gaussian kernels are used in transformations to project data over a large area.
• Linear classifiers are used to separate the classes from the changed space by expanding the margin.
• Few points, referred to as support vectors, can be used to locate the hyperplanes.

6. RESULTS AND DISCUSSION

6.1. Selection of the Overall Model

To enhance prediction accuracy, a method is employed that evaluates the outcomes of different machine learning algorithms. The method involves utilizing an approach, where a combination of learning models are trained to build a robust predictive model. Four well-known supervised algorithms—K-NN, Decision Tree, SVM, and Logistic Regression—are selected for in our study. Python serves as the language of choice for implementing these algorithms.

6.2. Pseudo Code Algorithm: Detection of Fake News

In this subsection, we outline the sequential steps undertaken:

- Input: Utilize a dataset containing fake news.
- Output: Produce Classifier Results.
- Cleaning dataset: Eliminate the Id column from datasets.

If any dataset attribute value is 'NULL', proceed to:

- Remove the corresponding row from the dataset.
- Extract content by selecting the 'TEXT' attribute.
- Perform necessary preprocessing steps to clean the text content.
- Tokenize the cleaned text content.
- Convert tokenized text content into a numerical representation through vectorization, suitable for training.
- Split the dataset into two subsets: the first for training and the second for testing.
- Employ the selected algorithm to train a classifier model using the training set.
- Evaluate the trained model's performance using the test set.
- Acquire the classifier results.
- Utilize the classifier for predictions if required.

6.3. Results of the Experiment

The experiment conducted on the Fake News datasets utilized Python 3.7 as the programming language. In this study, 70% of the records were allocated for training the classifiers, while the remaining 30% were reserved for testing the classifiers using k-fold cross-validation. The subsequent section will delve into the outcomes derived from analyzing the fake news datasets.

Moving forward, the implementation of the models will be carried out to obtain results, aiming to assess the effectiveness of the models based on calculated parameters. Various performance measures will be employed to evaluate the performance of the machine learning models, including F1-score, Precision, and Recall. These metrics serve as benchmarks for assessing the efficacy of the models.

For evaluation purposes, four key metrics are utilized and defined in Table 2.

Table 2. Metric abbreviations

Case name	Abbreviation	Description
True positive	TP	Indicates values correctly classified as positive
False positive	FP	Indicates values classified as positive in error
True negative	TN	Indicate correctly classified negative values
False negative	FN	Indicates values classified as negative in error

In a classification problem, a machine learning model should learn how to predict an input class, which implies the existence of a reference frame. Then, the computation of the confusion matrix requires the use of two data sets. The first set consists of a test dataset, also called test dataset, the second set serves as a repository and corresponds to the validation dataset.

1. Confusion matrix: is a metric called also contingency table. It is used to display the results of a predictive analysis. But in terms of our study with machine learning, it is used to investigate the performance of a model during a classification or prediction task. Among other things, the confusion matrix compares the real values to the values predicted by the model.

Table 3. Confusion matrix

		Predicted class	
		Positive	Negative
Actual class	Positive	FN	TP
	Negative	TN	FP

2. Accuracy tells us how well our model performs with new data. It is the percentage of correctly anticipated results. A high accuracy for the model can be an indication of a good model that can perform well with real-time data and can be deployed in practice. Accuracy is defined with the Equation (1):

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

3. Recall tells us the percentage of the positive examples that a model has automatically ranked among all positive examples. It can also be called success rate. Recall is defined with the Equation (2):

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

4. The ratio of true positives to total positive observations is determined by precision, but it is primarily concerned with data that has been labeled as positive by the model. The definition is provided in Equation (3):

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

5. F-Measure known as F1-Score, allows to determine precision and recall simultaneously thanks to the harmonic mean. F-Measure can be defined in Equation (4):

$$F1\text{-}Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

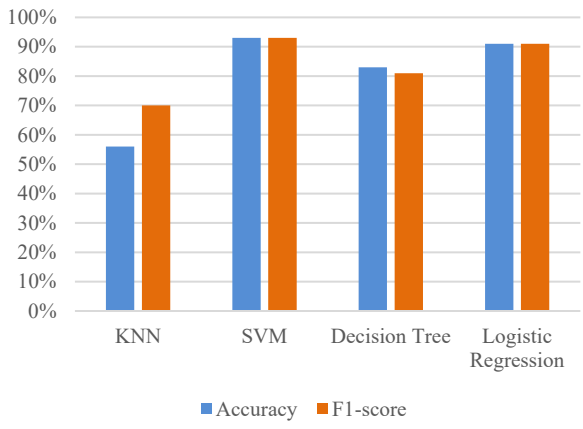

Figure 4. Accuracy vs F1-score for the different algorithms.

Table 4. Overall comparison of the accuracy on our dataset

Algorithms	Accuracy	Recall	Precision	F1-score
KNN	56.12	99.02	54.62	70.40
SVM	92.9	92.54	92.84	92.68
Decision Tree	82.87	84.07	79.83	81.89
Logistic Regression	91.55	90.18	92.68	91.41

Figure 4 and Table 4 depict results that help define the algorithm and provide a reliable prediction.

Based on the results presented in Figure 4 and Table 4, KNN performs relatively poorly compared to the other algorithms. While it achieves a high recall rate (99.02%), indicating it effectively identifies positive cases, the precision is quite low (54.62%), meaning it misclassifies a significant number of negative cases as positive. Consequently, the F1-score and accuracy are also relatively low, indicating overall poor performance, compared to SVM which performs very well across all metrics. It achieves high precision, recall, F1-score, and accuracy, indicating it effectively identifies both positive and negative cases with minimal misclassification. Additionally, Decision Tree performs reasonably well, with balanced precision and recall rates. However, it lags slightly behind SVM in terms of accuracy and F1-score, indicating it may misclassify some instances compared to SVM. Similar to SVM, Logistic Regression also performs well, with high precision, recall, F1-score, and accuracy. It achieves a good balance between identifying positive cases and avoiding false positives.

In summary, SVM and Logistic Regression outperform KNN and Decision Tree in terms of all metrics, indicating their superiority for the given classification task. SVM particularly stands out with consistently high performance across all metrics.

In the future, we aim to test others algorithms on our datasets to determine its robustness, also we plan to test our dataset in other algorithms for unsupervised learning and semi supervised learning, in addition, we intend to conduct a survey to evaluate the results.

7. CONCLUSIONS

The detection of fake news has evolved as an important challenge in today's digital era, where misinformation spreads rapidly across various online platforms. From a data mining perspective, the study has explored the effectiveness of different machine learning algorithms in identifying fake news.

Through comprehensive evaluation using metrics such as F1-score, precision, recall, and accuracy, it's evident that certain algorithms, notably Support Vector Machine (SVM) and Logistic Regression, exhibit robust performance. These algorithms demonstrate higher precision, recall, and accuracy compared to K-Nearest Neighbors (KNN) and Decision Tree algorithms. In our paper, we proposed a supervised machine learning model for detecting false news. Our model demonstrated superior performance compared to existing approaches. Testing the model on Fake news datasets yielded accuracies ranging from 56.12% to 92.9%, and achieving 82.87% with the Decision Tree model.

Overall, the study underscores the potential of data mining approaches in combating the proliferation of fake news, offering valuable insights for the development of more effective detection mechanisms to safeguard the integrity of online information dissemination.

Moving forward, future work will focus on expanding the scope of our study. Which will involve incorporating additional datasets to further enhance the robustness and generalizability of the model. Additionally, we plan to explore the efficacy of alternative classifiers to compare and potentially improve upon the performance of our current model. This iterative approach aims to continually refine and advance the capabilities of false news detection systems.

REFERENCES

[1] C. Sindermann, H.S. Schmitt, D. Rozgonjuk, J.D. Elhai, C. Montag, "The Evaluation of Fake and True News: on the Role of Intelligence, Personality, Interpersonal Trust, Ideological Attitudes, and News Consumption", Heliyon, Vol. 7, No. 3, p. e06503, 2021.

[2] O. Dardary, M. Tridane, S. Belaaouad, "Pedagogical Practices and Academic Achievement in the Era of Global Pandemics", International Journal on Technical and Physical Problems of Engineering (IJTPE), Iss. 54, Vol. 15, No. 3, pp. 81-86, March 2023.

[3] J.C. Reis, A. Correia, F. Murai, A. Veloso, F. Benevenuto, "Supervised Learning for Fake News Detection", IEEE Intelligent Systems, Vol. 34, No. 2, pp. 76-81, April 2019.

[4] M.H. Goldani, S. Momtazi, R. Safabakhsh, "Detecting Fake News with Capsule Neural Networks", Applied Soft Computing, Vol. 101, p. 106991, 2021.

[5] Y. Liu, Y.F.B. Wu, "FNED: A Deep Network for Fake News Early Detection on Social Media", ACM Transactions on Information Systems, Vol. 38, No. 3, pp. 1-33, 2020.

[6] M.D. Vicario, W. Quattrociocchi, A. Scala, F. Zollo, "Polarization and Fake News: Early Warning of Potential Misinformation Targets", ACM Transactions on the Web, Vol. 13, No. 2, pp. 1-22, 2019.

[7] M.Z. Asghar, A. Habib, A. Habib, A. Khan, R. Ali, A. Khattak, "Exploring Deep Neural Networks for Rumor Detection", Journal of Ambient Intelligence and Humanized Computing, Vol. 12, pp. 4315-4333, 2021.

[8] A. Vereshchaka, S. Cosimini, W. Dong, "Analyzing and Distinguishing Fake and Real News to Mitigate the Problem of Disinformation", Computational and Mathematical Organization Theory, Vol. 26, pp. 350-364, 2020.

[9] A.P.H.A. Faustini, T.F. Covoes, "Fake News Detection in Multiple Platforms and Languages", Expert Systems with Applications, Vol. 158, pp. 113503, 2020.

[10] S. Rastogi, D. Bansal, "A Review on Fake News Detection 3T's: Typology, Time of Detection, Taxonomies", International Journal of Information Security, Vol. 22, pp. 177-212, 2023.

[11] R.K. Kaliyar, A. Goswami, P. Narang, "DeepFakE: Improving Fake News Detection Using Tensor Decomposition-Based Deep Neural Network", The

Journal of Supercomputing, Vol. 158, No. 77, pp.1015-1037, 2021.
[12] S.S. Jadhav, S.D. Thepade, "Fake News Identification and Classification Using DSSM and Improved Recurrent Neural Network Classifier", Applied Artificial Intelligence, Vol. 33, No. 12, pp. 1058-1068, 2019.
[13] M. El Mohadab, B. Bouikhalene, S. Safi, "Automatic CV Processing for Scientific Research Using Data Mining Algorithm", Journal of King Saud University - Computer and Information Sciences, Vol. 32, pp. 561-567, 2020.
[14] M. El Mohadab, B. Bouikhalene, S. Safi, "Predicting Rank for Scientific Research Papers Using Supervised Learning", Applied Computing and Informatics, Vol. 15, pp. 182-190, 2019.
[15] A.O. Mohd, H.M Noor, "A Survey on Unsupervised Learning for Wearable Sensor-Based Activity Recognition", Applied Soft Computing, Vol. 127, pp. 109363, 2022.
[16] A. Ligtharta, C. Catalb, B.Tekinerdogan, "Analyzing the Effectiveness of Semi-Supervised Learning Approaches for Opinion Spam Classification", Applied Soft Computing, Vol. 101, pp. 107023, 2021.
[17] A.E. Pakzad, R.M. Manuel, J.S. Uy, X.F. Asuncion, J.V. Ligayo, L. Materum, "Reinforcement Learning-Based Television White Space Database Classification", Baghdad Science Journal, Vol. 18, No. 2, p. 0947, June 2021.
[18] https://www.kaggle.com.
[19] O. El Melhaoui, S. Benchaou, R. Zarrouk, A. Bouzidi, "Enhancing Signature Recognition Performance Through Convolutional Neural Network and K-Nearest Neighbors", International Journal on Technical and Physical Problems of Engineering (IJTPE), Iss. 57, Vol. 15, No. 4, pp. 443-449, December 2023.
[20] E. Dumitrescua, S. Hueb, C. Hurlinb, S. Tokpavi, "Machine Learning for Credit Scoring: Improving Logistic Regression with Non-Linear Decision-Tree Effects", European Journal of Operational Research, Vol. 297, No. 3, pp. 1178-1192, 2022.
[21] V. Yousefi, S. Kheiri, S. Rajebi, "Evaluation of K-Nearest Neighbor, Bayesian, Perceptron, RBF and SVM Neural Networks in Diagnosis of Dermatology Disease", International Journal on Technical and Physical Problems of Engineering (IJTPE), Iss. 42, Vol. 12, No. 1, pp. 114-120, March 2020.
[22] J. Alcaraza, M. Labbeb, M. Landetea, "Support Vector Machine with Feature Selection: A Multiobjective Approach", Expert Systems with Applications, Vol. 204, p. 117485, 2022.

BIOGRAPHIES

Name: **Mohamed**
Surname: **El Mohadab**
Birthday: 09.10.1990
Birthplace: Beni Mellal, Morocco
Bachelor: Networking and Telecommunication, Computer Science Department, Polydisciplinary Faculty, Moulay Ismail University, Errachidia, Morocco, 2013
Master: Computer Science, Computer Science Department, Polydisciplinary Faculty, Sulatn Moulay Slimane University, Beni Mellal, Morocco, 2015
Doctorate: Computer Science, Polydisciplinary Faculty, Sulatn Moulay Slimane University, Beni Mellal, Morocco, 2020
The Last Scientific Position: Assist. Prof., Computer Science department, Faculty of Science, Chouaib DoukKali University, El Jadida, Morocco, Since 2020
Research Interests: Data Mining, Machine Learning, Information System
Scientific Publications: 9 Papers, 2 Book Chapters
Scientific Memberships: LAROSERI, LIMATI, LGS

Name: **Belaid**
Surname: **Bouikhalene**
Birthday: 01.01.1970
Birthplace: Beni Mellal, Morocco
Bachelor: Physics, Electronic, Sultan Moulay Slimane University, Beni Mellal, Morocco, 1992
Master: Physics and Electronic, Sultan Moulay Slimane University, Beni Mellal, Morocco, 1994
Doctorate: Computer Sciences, Cady Ayyad University, Marrakech, Morocco, 2006
The Last Scientific Position: Prof., Mathematics and Computer Science Department, Polydisciplinary Faculty, University of Sultan Moulay Slimane, Beni Mellal, Morocco, Since 2007
Research Interests: Data Mining, Machine Learning, Mathematics, Information System
Scientific Publications: 173 Papers, 2 Books, 11 Chapters
Scientific Memberships: LIMATI

Name: **Said**
Surname: **Safi**
Birthday: 01.01.1973
Birthplace: Beni Mellal, Morocco
Bachelor: Physics, Electronic, Cady Ayyad University Marrakech, Morocco, 1995
Master: Physics and Electronic, Chouaib Doukkali University, El Jadida, Morocco, 1997
Doctorate: Telecommunication and Informatics, Cady Ayyad University, Marrakech, Morocco, 2008
The Last Scientific Position: Prof., Mathematics and Computer Science Department, Polydisciplinary Faculty, University of Sultan Moulay Slimane, Beni Mellal, Morocco, Since 2008
Research Interests: Data Mining, Telecommunication, Information System
Scientific Publications: 170 Papers, 1 Book, 13 Chapters
Scientific Memberships: LIMATI