Journal of Information Systems Engineering and Management
2025, 10(58s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Enhancement of Rumors and Fake News Predictions Based on Machine Learning Techniques

Manahil Zayno, Abdulkareem Merhej Radhi
[1] Department of Computer Engineering, College of Engineering, Al-Nahrain University, Jadriya, Baghdad, Iraq
[2] Department of Computer Science, College of Science, Al-Nahrain University, Jadriya, Baghdad, Iraq

ARTICLE INFO

Received: 30 Dec 2024

Revised: 19 Feb 2025

Accepted: 27 Feb 2025

ABSTRACT

Rumor's propagation is accelerating along with the rapid rise in popularity of social networks. Any unverefied statements that originate from one or more sources and subsequently spread throughout the meta-networks are referred to be rumors. Because rumors could occasionally be harmful, especially when it comes to social and political concerns that have a greater influence on people's lives, it is imperative to find strategies for detecting rumors as soon as feasible. This study is aimed at developing a classifier for Rumor Detection using Machine Learning (ML) algorithms and Natural Language Processing (NLP) tools. Four preprocessing types are utilized by the system. The most widely utilized feature extraction (FE) approach is the Term Frequency-Inverse Document Frequency (TF-IDF) technique, which comes next. Rumors are detected using Stochastic Gradient Descent (SGD) approach and five classification algorithms: Naive-Bayes (NB), Random Forest (RF), K-Nearest Neighbor (K-NN), Logistic Regression (LR), and Decision Tree (DT). The best accurate method for classifying rumor data was determined by comparing these six methods. Additionally, recall, precision, F1-score, and accuracy measurements have been used to evaluate the classification algorithms' results based on performance metrics. Results indicated that RF applied to three datasets produced the maximum accuracy, with 93% accuracy in the SNOPES dataset and 93% accuracy in the unbalanced dataset. The SGD approach yielded 96% accuracy for the balanced dataset Faknews

Introduction: Nowadays, social networks are with no doubt the most important for accessing information. Due to the large number of social network relations and the vast amount of data, scientists are working to overcome a number of obstacles resulting from the rich relations that social networks contain [Alatas and Altay, 2018]. Although social networks are an excellent to obtain news, they have become a powerful tool for manipulating people and a source of rumors on any topic. Rumor is a social problem with numerous negative reflections. Positive relations could deteriorate and the company's and individuals' reputations could be harmed. Serious financial harm could be done to individuals, businesses, families, and even entire nations. People have lost a lot of as a result of gossip events, which have become a major issue in social networks recently. For instance, rumors have a detrimental effect on the political landscape, the economy, and the stability of society. A rumor about "shootouts and kidnappings through drug gangs near schools in Veracruz" surfaced on Twitter and Facebook on August 25, 2015. There have been 26 car accidents in the city as a result of people leaving their cars in the middle of the road and hurrying to pick up their kids from school [Tang]. The necessity of the automatic prediction of authenticity of content on social media has been shown by this case of false rumors. A lot of research into creating systems that can identify rumors on their own depends on AI tools like NLP and ML..

Objectives: This paper aims to design and implement a Rumor Detection System with the use of ML and NLP. The system is designed for accurately classifying posts on Facebook and Twitter as either truths or rumors. It uses different ML algorithms, such as RF, NB, LR, K-NN, DT, and the SGD. The aim is achieving high accuracy, aiding opinion analysts in detecting and classifying rumors across diverse datasets.

Methods: At varius vel pharetra vel turpis nunc eget lorem. Feugiat scelerisque varius morbi enim nunc. Cras semper auctor neque vitae tempus quam pellentesque nec. Faucibus purus in massa tempor nec feugiat nisl. Congue nisi vitae suscipit tellus mauris a. Est sit amet facilisis

Journal of Information Systems Engineering and Management
2025, 10(58s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

magna etiam tempor. Dictum varius duis at consectetur. Purus semper eget duis at tellus at urna. Ipsum consequat nisl vel pretium. Viverra maecenas accumsan lacus vel facilisis volutpat est. Bibendum arcu vitae elementum curabitur vitae nunc sed. Nisl tincidunt eget nullam non nisi est. Ac turpis egestas integer eget aliquet nibh praesent.

Results: Egestas diam in arcu cursus euismod quis viverra nibh. Convallis aenean et tortor at risus viverra. Sit amet justo donec enim diam. Sem et tortor consequat id. Purus gravida quis blandit turpis. Consectetur adipiscing elit duis tristique sollicitudin nibh sit amet commodo. Eget duis at tellus at urna condimentum mattis pellentesque. Auctor elit sed vulputate mi sit amet. Consequat ac felis donec et. In dictum non consectetur a erat nam at lectus. Dui vivamus arcu felis bibendum ut tristique. Lacinia quis vel eros donec ac. Ac turpis egestas maecenas pharetra convallis posuere morbi leo. Tortor id aliquet lectus proin.

Conclusions: Mi tempus imperdiet nulla malesuada. Magna fermentum iaculis eu non diam phasellus vestibulum. Consectetur adipiscing elit duis tristique sollicitudin nibh sit amet commodo. Elit scelerisque mauris pellentesque pulvinar. Et malesuada fames ac turpis egestas maecenas pharetra convallis posuere. Elementum integer enim neque volutpat ac tincidunt vitae semper.

Keywords: fermentum, condimentum, tristique, viverra

INTRODUCTION

Nowadays, social networks are with no doubt the most important for accessing information. Due to the large number of social network relations and the vast amount of data, scientists are working to overcome a number of obstacles resulting from the rich relations that social networks contain [1]. Although social networks are an excellent to obtain news, they have become a powerful tool for manipulating people and a source of rumors on any topic. Rumor is a social problem with numerous negative reflections. Positive relations could deteriorate and the company's and individuals' reputations could be harmed. Serious financial harm could be done to individuals, businesses, families, and even entire nations. People have lost a lot of as a result of gossip events, which have become a major issue in social networks recently. For instance, rumors have a detrimental effect on the political landscape, the economy, and the stability of society. A rumor about "shootouts and kidnappings through drug gangs near schools in Veracruz" surfaced on Twitter and Facebook on August 25, 2015. There have been 26 car accidents in the city as a result of people leaving their cars in the middle of the road and hurrying to pick up their kids from school [2]. The necessity of the automatic prediction of authenticity of content on social media has been shown by this case of false rumors. A lot of research into creating systems that can identify rumors on their own depends on AI tools like NLP and ML.

Objectives

This paper aims to design and implement a Rumor Detection System with the use of ML and NLP. The system is designed for accurately classifying posts on Facebook and Twitter as either truths or rumors. It uses different ML algorithms, such as RF, NB, LR, K-NN, DT, and the SGD. The aim is achieving high accuracy, aiding opinion analysts in detecting and classifying rumors across diverse datasets.

The Proposed System Architecture

To automatically detect rumors from text content, There are several stages related to the architecture of the suggested system. Classification methods that are applied are represented on the fifth stage. This encompasses the application of NB, DT, RF, KNN, LR, and SGD. The final stage is modeled performance evaluation, different metrics such as (accuracy, recall, and precision) have been used as described in Figure (1).

Journal of Information Systems Engineering and Management
2025, 10(58s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

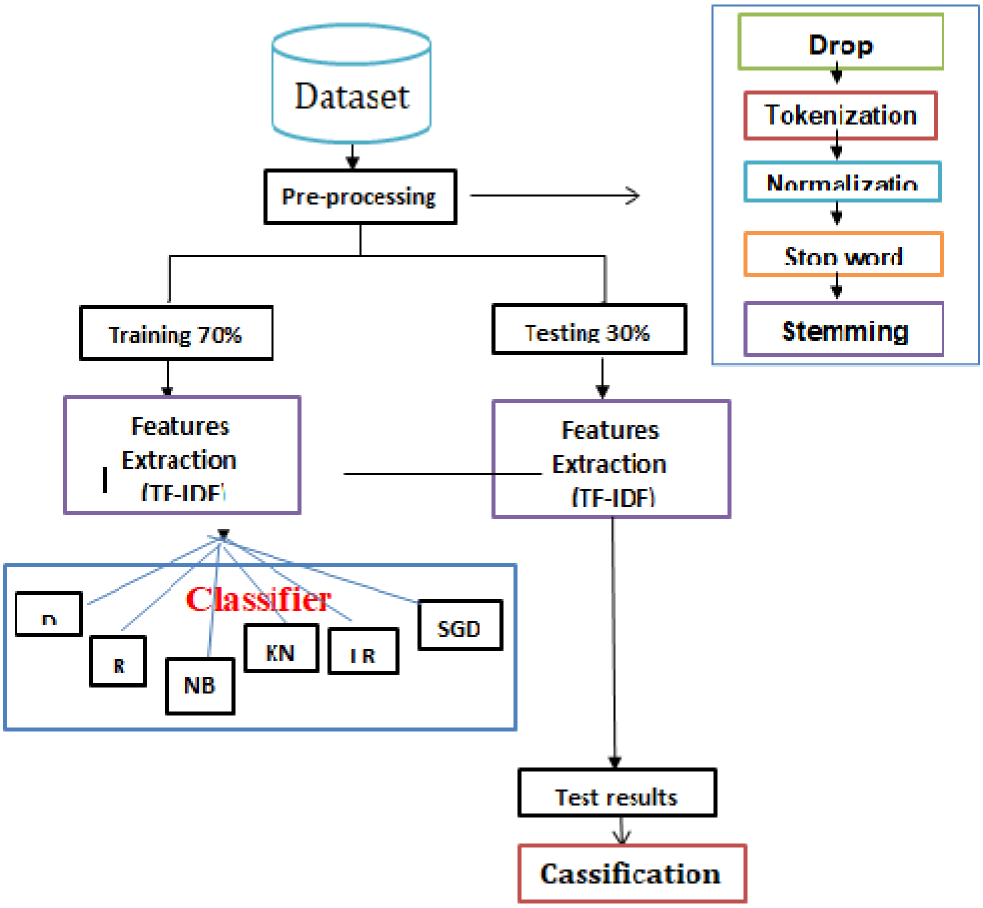


Figure (1) Block Diagram the Proposed System

5.1 Datasets

In the classification process, the datasets are crucial. This huge collection was gathered from kaggle.com. Dataset is utilized for training and testing the classifier. Two datasets have been acquired in the present thesis. A rumor dataset consisting of three files and a fake news dataset is used that is collected from kaggel.com.

Since the Rumor dataset has duplicates so the first stage in preprocessing is removing duplicates. After removing duplicates a rumor faced an unbalanced issue. The smote and Adaptive Synthetic Sampling methods were applied to solve this issue as explained deeply in further steps.

5.2 Preprocessing stages

This phase is essential to any NLP system since the characters, words, and units detected at this level serve as the basic units for all further processing stages. The pre-processing steps that are the same for the testing and training phases are demonstrated in the portion that is being presented. The fact that this stage arranges the data to make Rumor Detection easier is a significant advantage.

5.2.2 Tokenization

Tokenization is essential in NLP. At this point, the text document is divided into tokens, with each word being separated from the others by a space. Tokens might be numbers, words, or symbols. As seen in Fig (2), each document was divided into tokens based on whitespaces.

Journal of Information Systems Engineering and Management
2025, 10(58s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

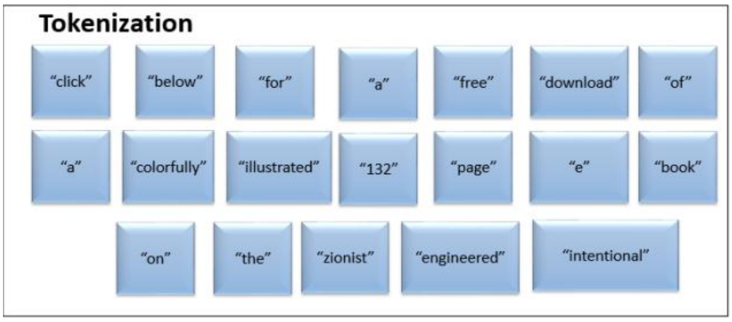


Figure (2) Example of Tokenization process

5.2.2 Normalization

This is the process used to unify the various forms of of the same letter through converting each character into upper or lower cases, also all of the numbers and symbols are deleted, like *, #, $, !, ?, @, [ ], { }, %, &, (),! =, +, <, > =, -, _,: -, ", '\ and /. Every word is converted into Lower Case (LC).

5.2.3 Remove Stop Words

Stop words represent the frequently used words that can be defined as any word which is irrelevant in classification or does not have any distinguishing importance. Eliminating stop words reduces dimensionality because they are longer than 400 words. Consequently, keywords that were left in the Rumor dataset may be easily identified using automatic feature extraction techniques.

5.3 Features Extraction

For providing the raw text data to a statistical or ML algorithm, the FE process is carried out for extracting meaningful features or attributes. Since numeric vectors from raw text tokens are frequently the outcome of this technique, it was called as vectorization. This is due to the fact that conventional algorithms cannot directly work with raw textual data; instead, they operate on numerical vectors. This FE contains a bag of words (Bow) and TF-IDF. While TF-IDF considers inverse document frequency and term frequency when weighing every term, the Bow displays frequency of occurrence of a word or group of words in some text [9]. But it is predicated on the Bow method. The TF-IDF paradigm, which will be covered in more detail, is thus adopted in this study.

5.4. Term Frequency-Inverse Document Frequency

A popular technique for obtaining feature entries in the text classification process is TF-IDF algorithm, which is straightforward and incredibly effective. A statistical technique for determining a word's significance for one document or some corpuses is called TF-IDF. A word's importance rises in direct proportion to how many times it occurs in the document, yet it falls in direct proportion to how frequently it occurs in the corpus. The number of times a word somes up in a file is known as its term frequency, or TF [10].

A) TF-IDF

On the contrary with reverse proportion with respect to the same feature on the documents in a training dataset, TF-IDF focuses on defining relative frequency with regard to the characteristics that are present in a particular document. The purpose of this computation is to ascertain the significance of the designated features in a given document. Eq2.3 will be used for the calculation of TF-IDF as follows:

$$\mathrm{TF} - \mathrm{IDF} = \mathrm{TF_i} * (\log_2(\frac{\mathrm{N}}{\mathrm{N_i}})) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

The number of times a word "i" has appeared in a document is indicated by the symbol "TF." N: stands for the total number of documents in a set of documents. Ni is a measure of how many times a word appeared in a set of documents [11].

Journal of Information Systems Engineering and Management
2025, 10(58s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ Research Article

B) Entropy

Entropy measures the disorder or unpredictability of the system. It is described as shown in Eq. (2):

$$H(p) = -\sum_{i=1}^{c} P_i \log_2 P_i \text{ .... } (2)$$

C) Information Gain (IG)

For figuring out the optimal attributes to select for each decision node in a tree. The attribute that best divides them into homogeneous subsets is the best attribute. More specifically, the equation (3) displays the attribute A's Gain (S, A), depending on a set of samples S.

$$\text{Gain (S, A) = Entropy (S)-} \sum_{i=1}^{k} \frac{|S_i|}{|S|} \text{ Entropy } (Sv_i) \text{ ....} (3)$$

Where attribute A has a set of values [12],, and the number of the examples within S with a value vi is |Svi| [13].

DT has many pros likepresented in [14]:

6. Evaluation Metrics Methods

Several metrics of evaluation, which includes precision, accuracy, recall, and F1-score, have been utilized for the assessment of each model. Because different metrics do not account for the same values, it is necessary to use multiple metrics [15]. These metrics allow the classifier's performance to be assessed from multiple perspectives and are frequently used in the ML community [16]. The confusion matrix, which is a matrix where test papers are distributed through the establishment of 2 classes as described in Table 1, must be calculated in order to calculate those measures. Below is a description of this table's key values.

Table (1) confusion matrix [Kadhim, 2019][17]

	Predicted Classes	
class Actual	Positives (P)	Negatives (N)
True (T)	True Positive (TP)	False Negative (FN)
False (F)	False Positive (FP)	True Negative (TN)

6.1 Accuracy Measure

One of the most commonly utilized metrics for classification performance is accuracy, which may be defined as the ratio of the correctly classified samples to all samples, as shown in equation (4) [18].

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \text{ .... } (4)$$

It will be feasible to compute TF-IDF after IDF and TF values had been determined. Algorithm (1) describes the steps of TF-IDF algorithm.

Algorithm(1) Features Extraction
Inputs: Pre-processing Dataset
Outputs: TF-IDF for every one of the words
Begin Step 1: For each word in dataset do Begin Construct a list of features from a set of tokens, each feature associated with its index.

Journal of Information Systems Engineering and Management
2025, 10(58s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

End
Step 2: Compute term frequency (TF) eq (2).
Calculate the inverse document frequency (IDF) with the use of eq (3). Compute TF-IDF by using eq (1).

Step3: Return a vector with one row for each Record t and one column for each of the features where every one of the features is related to its weights (TF-IDF)
End.

7.Oversampling Techniques

Although the dataset is imbalanced, the findings that were produced are good. To improve the results, oversampling was used to balance the dataset. One of the most popular methods for handling an imbalanced dataset is resampling it. Generally speaking, oversampling is better than undersampling methods. The reason is that undersampling tends to eliminate data points that might contain significant information. SMOTE and its associated equivalents are among the particular data augmentation oversampling techniques that are especially covered in this study.

8. SMOTE: Synthetic Minority versamplingTechnique

Synthetic samples are created for minority class with the use of SMOTE oversampling method. With the aid of the interpolation between positive instances that lay together, it is concentrated on feature space for the production of new instances. The solution for oversampling is provided in algorithm (2).

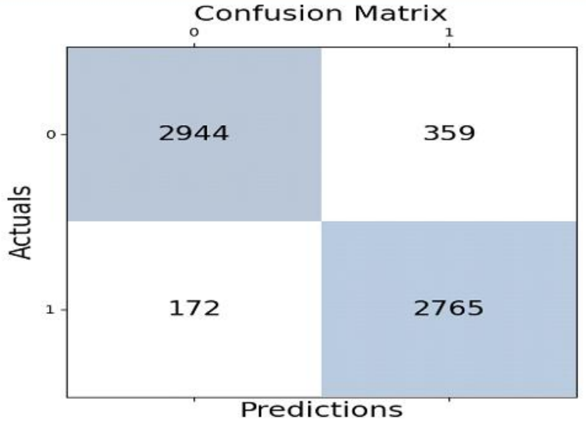
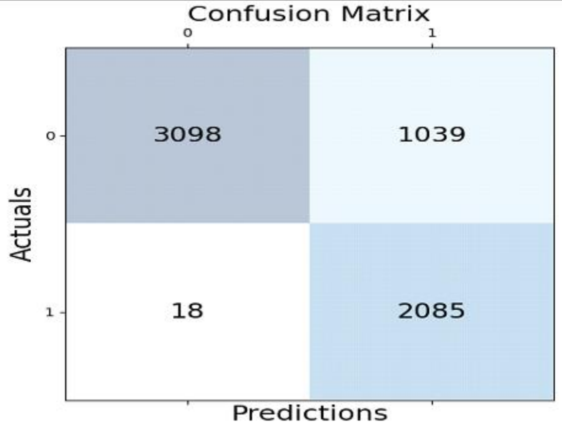
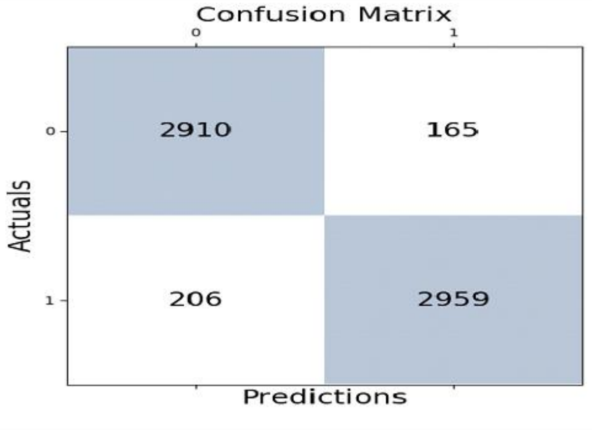
Algorithm (2): Overcoming Class Imbalance using SMOTE Techniques
Inputs: Imbalanced dataset
Outputs: balanced dataset
Begin Step 1: For I =1 to N// every oversampling observation. Begin Select a positive class instance randomly. End Step 2: KNN's (by default 5) for that instance has been obtained. // as descried in chapter2// Step 3: chose N of those K instances for interpolating new synthetic instances Step 4: compute any distance metric the difference in the distance between feature vector and its neighbours Step 5: multiply a difference is by any random value in (0,1] and is added to previous feature vector. End.

9.EXPERIMENTS AND EVALUATION OF RESULTS

The suggested approach's performance for solving the rumor detection problem is evaluated in this paper, and explains the results of the processes present that are utilized in the system such as preprocessing phase and feature extraction as their methods are mentioned in chapter three, as well as measuring the accuracy, and performance of the system for Rumor detection. In addition to that, the performance of other algorithms used for comparison purposes. In general, the stages of achievements which applied on fourths data naming Emergent dataset, Politifact dataset, and SNOPS dataset, Fake news dressed.

Journal of Information Systems Engineering and Management
2025, 10(58s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

<table>
<tr><th></th><th>Table (3) confusion matrix results for fake news datasetAlgorithm
Confusion Matrix</th></tr>
<tr><td>RF</td><td>
</td></tr>
<tr><td>NB</td><td>
</td></tr>
<tr><td>LR</td><td>
</td></tr>
</table>

The suggested system is applied to all datasets, including balanced and unbalanced datasets. We got an excellent result from SNOPS dataset in the imbalanced dataset, and a good result from Fake news data-set in the balanced dataset; we'll go through both results in detail in the next sections.

Journal of Information Systems Engineering and Management
2025, 10(58s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

10. Results of the SGD method for fake news dataset

For the iterative solution of the SGD the results of confusion matrix for the fake news dataset as can be seen in figure 3.

SGD

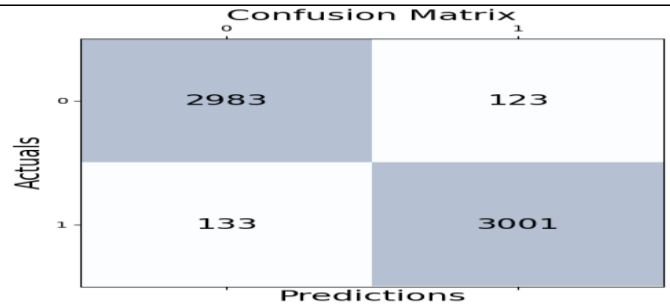


Figure (3) SGD confusion matrix results for fake news dataset

Table 4 shows results of Precision using equation the previous equations

Table 4 shows Results of Unbalanced data

RF		precision	recall	f1-score	support
	0	0.93	0.88	0.91	3226
	1	0.88	0.93	0.91	3014
	accuracy			0.91	6240
	macro avg	0.91	0.91	0.91	6240
	weighted avg	0.91	0.91	0.91	6240

Journal of Information Systems Engineering and Management
2025, 10(58s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

		precision	recall	f1-score	support
RF	FALSE	1.00	0.55	0.54	2254
	TRUE	0.82	1.00	0.90	253
	legend	0.78	1.00	0.87	13
	mfalse	0.82	0.99	0.90	425
	mixture	0.82	1.00	0.90	352
	mtrue	0.81	1.00	0.89	67
	undetermined	0.83	1.00	0.91	29
	accuracy			0.93	3433
	macro avg	0.84	0.98	0.90	3433
	weighted avg	0.94	0.93	0.93	3433
NB	FALSE	1.00	0.60	0.75	3429
	TRUE	0.00	0.00	0.00	0
	legend	0.00	0.00	0.00	0
	mfalse	0.01	1.00	0.01	3
	mixture	0.00	1.00	0.00	1
	mtrue	0.00	0.00	0.00	0
	undetermined	0.00	0.00	0.00	0
	accuracy			0.60	3433
	macro avg	0.14	0.37	0.11	3433
	weighted avg	1.00	0.60	0.75	3433
LR	FALSE	1.00	0.61	0.76	3338
	TRUE	0.02	1.00	0.04	6
	legend	0.00	0.00	0.00	0
	mfalse	0.12	0.97	0.21	62
	mixture	0.06	0.96	0.11	27
	mtrue	0.00	0.00	0.00	0
	undetermined	0.00	0.00	0.00	0
	accuracy			0.62	3433
	macro avg	0.17	0.51	0.16	3433
	weighted avg	0.97	0.62	0.74	3433
KNN	FALSE	0.95	0.65	0.77	2999
	TRUE	0.16	0.51	0.24	94
	legend	0.00	0.00	0.00	0
	mfalse	0.19	0.53	0.28	185
	mixture	0.16	0.46	0.24	149
	mtrue	0.01	0.20	0.02	5
	undetermined	0.03	1.00	0.06	1
	accuracy			0.63	3433
	macro avg	0.21	0.48	0.23	3433
	weighted avg	0.85	0.63	0.70	3433
DT	FALSE	[illegible]	[illegible]	[illegible]	[illegible]
	TRUE	[illegible]	[illegible]	[illegible]	[illegible]
	legend	[illegible]	[illegible]	[illegible]	[illegible]
	mfalse	[illegible]	[illegible]	[illegible]	[illegible]
	mixture	[illegible]	[illegible]	[illegible]	[illegible]
	mtrue	[illegible]	[illegible]	[illegible]	[illegible]
	undetermined	[illegible]	[illegible]	[illegible]	[illegible]
	accuracy			[illegible]	[illegible]
	macro avg	[illegible]	[illegible]	[illegible]	[illegible]
	weighted avg	[illegible]	[illegible]	[illegible]	[illegible]

As shown in table (4) the best accuracy value overall classification algorithms are obtained with Snopes data-set and TF-IDF feature extraction methods, the best accuracy value of the RF classifier =0.93, the best accuracy value of the NB classifier =0.60, the best accuracy value of the LR classifier =0.62, and the best accuracy value of the KNN =0.63, DT = 0.90 and SGD = 0.90. Table (4) illustrates all results of use evaluation metrics for every one of the classifiers about the Fake news data-set.

Table (5) All Results Fake News Metrics for each classifier about Fake News Dataset

		precision	recall	f1-score	support
NB	0	1.00	0.72	0.83	4241
	1	0.62	0.99	0.77	1999
	accuracy			0.81	6240
	macro avg	0.81	0.86	0.80	6240
	weighted avg	0.88	0.81	0.81	6240
LR	0	0.93	0.95	0.94	3024
	1	0.95	0.94	0.94	3216
	accuracy			0.94	6240
	macro avg	0.94	0.94	0.94	6240
	weighted avg	0.94	0.94	0.94	6240
KNN	0	0.28	0.96	0.43	883
	1	0.99	0.59	0.74	5357
	accuracy			0.64	6240
	macro avg	0.63	0.77	0.58	6240
	weighted avg	0.89	0.64	0.69	6240
DT	0	0.87	0.87	0.87	3064
	1	0.88	0.88	0.88	3176
	accuracy			0.87	6240
	macro avg	0.87	0.87	0.87	6240
	weighted avg	0.87	0.87	0.87	6240

11. Comparing the Proposed Model's Results With Other Related Works

Journal of Information Systems Engineering and Management
2025, 10(58s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

According to this section, the comparison of the proposed models of Rumor and Fake news with previous related work has been done. Tables (5) and (6) illustrates the comparison between proposed work and related works on the Rumor and Fake news dataset

Table (5) Comparing the result with other related work for Rumor

Author(s), Year	The algorithm for the classifier	Accuracy %
Our proposed models, 2022 SNOPS dataset	Decision Tree RF KNN LR Naïve Bayesian Stochastic Gradient Descent	the best accuracy 93 % of the Random forest
(Habib, Akbar, Asghar, Khattak, Ali, & Batool, 2018)SNOP	Logistic Regression, Naïve Bayesian SVMs K-Nearest Neighbors	the best accuracy 72.43 %. In the Naïve Bayesian (NB),

In comparison with other models ‘ Habib et al. “ Our model, based on different approaches that used cross validation in, 10 folds, then get high accuracy = 0.93 in RandomFores classifier**.**

Table (6) Comparing the result with other related work for Fake news

Author(s), Year	The algorithm for the classifier	Accuracy
Our proposed models, 2022 Fake news dataset =20387	Naïve Bayesian Decision Tree RF KNN LR Stochastic Gradient Descent	the best accuracy96 % in the Stochastic Gradient Descent
(Adiba, Islam, aiser, Mahmud, & ahman, 2020) Fake news dataset =20387	Naive Bayes classifier	the best accuracy92% in the Naive Bayes

12. Conclusions

The detection of rumors in social media by supervised learning algorithms is the focus of this research. We tested many ML classifiers, and based upon outcomes, our model came to the following conclusion: -

i. The experimental evaluations were performed on two type of dataset which are balanced and unbalanced dataset, in our model we apply tow techniques to balance dataset.

ii. In the present study, the cross validation method is used to achieve accurate results within 10 folds

iii. The best algorithm used is Random Forest classified In an unbalanced dataset the SNOPS dataset archived the high accuracy: it outperformed all other classifiers with an accuracy of 93%.

iv. Another dataset that was used to experience the system is the balanced fake news dataset. The maximum accuracy score was in the SGD method 96%. The achieved result is better than the listed related works.

Journal of Information Systems Engineering and Management
2025, 10(58s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Refrences

1. ALTAY, E. V. & ALATAS, B. Detection of cyberbullying in social networks using machine learning methods. 2018 International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism (IBIGDELFT), 2018. IEEE, 87-91.

2. TANG, T. Course Report for Detecting Rumors from Microblogs with Recurrent Neural Networks.

3. HABIB, A., AKBAR, S., ASGHAR, M. Z., KHATTAK, A. M., ALI, R. & BATOOL, U. Rumor detection in business reviews using supervised machine learning. 2018 5th International Conference on Behavioral, Economic, and Socio-Cultural Computing (BESC), 2018. IEEE, 233-237.

4. JEHAD, R. & YOUSIF, S. A. 2020. Fake News Classification Using Random Forest and Decision Tree (J48). Al-Nahrain Journal of Science, 23, 49-55.

5. DUEBY, A.K., SINGHAL, A.& GUPTA, S.2020, Rumor Detection System using Machine Learning. International Research Journal of Engineering and Technology (IRJET). Volume:07 Issue: 05 1 May.

6. ADIBA, F. I., ISLAM, T., KAISER, M. S., MAHMUD, M. & RAHMAN, M. A. 2020. Effect of corpora on classification of fake news using naive Bayes classifier. International Journal of Automation, Artificial Intelligence and Machine Learning, 1, 80-92.

7. FRIGGERI, A., ADAMIC, L., ECKLES, D. & CHENG, J. Rumor cascades. Proceedings of the International AAAI Conference on Web and Social Media, 2014.

8. CAO, J., GUO, J., LI, X., JIN, Z., GUO, H. & LI, J. 2018. Automatic rumor detection on microblogs: A survey. arXiv preprint arXiv:1807.03505.

9. SARKAR, D. 2016. Text Analytics with python, Springer.Book.

10. LIU, C.-Z., SHENG, Y.-X., WEI, Z.-Q. & YANG, Y.-Q. Research of text classification based on improved TF-IDF algorithm. 2018 IEEE International Conference of Intelligent Robotic and Control Engineering (IRCE), 2018. IEEE, 218-222.

11. ALMUSADDAR, M. Y. 2014. Improving Arabic Light Stemming in Information Retrieval Systems.

12. ZHANG, S., LI, X., ZONG, M., ZHU, X. & CHENG, D. 2017. Learning k for knn classification. ACM Transactions on Intelligent Systems and Technology (TIST), 8,1-19.

13. DOUGHERTY, G. 2012. Pattern recognition and classification: an introduction, Springer Science & Business Media. Book.

14. CHARBUTY, B. & ABDULAZEEZ, A. 2021. Classification based on decision tree algorithm for machine learning. Journal of Applied Science and Technology Trends, 2, 20-28.

15. LORENT, S. 2019. Master thesis: Fake news detection using machine learning.

16. SHU, K., SLIVA, A., WANG, S., TANG, J. & LIU, H. 2017. Fake news detection on social media: A data mining perspective. ACM SIGKDD explorations newsletter, 19, 22-36.

17. KADHIM, A. I. 2019. Survey on supervised machine learning techniques for automatic text classification. Artificial Intelligence Review, 52, 273-292.

18. THARWAT, A. 2020. Classification assessment methods. Applied Computing and Informatics. Vol. 17 No. 1, 2021 pp. 168-192.