FAKE-NEWS DETECTION SYSTEM USING MACHINE-LEARNING ALGORITHMS FOR ARABIC-LANGUAGE CONTENT

MOUTAZ ALAZAB[1*], ALBARA AWAJAN[1], AMMAR ALAZAB[3], ANSAM KHREISAT[4], ABEER ALHYARI[2], REEM SAADEH[1]

[1]Intelligence System Department, Faculty of Artificial Intelligence, Al-Balqa Applied University
[2]Computer Engineering department, Faculty of Engineering, Al-Balqa Applied University
(m.alazab, a.awjan, a.alhyari, reemsdh)@bau.edu.jo
[3]School of IT & Engineering, Melbourne Institute of Technology
(aalazab)@mit.edu.au
[4] College of Business and Law, RMIT University
(ansam.khraisat)@ rmit.edu.au

ABSTRACT

Over the past decade, social media has become a dominant source of news and information. This has led to an increase in the number of groups and individuals spreading news through social media with no direct quality control or censorship of the content being distributed. A fake breaking-news headline can spread rapidly to millions of people and cause tremendous local and global problems. Because checking all information posted on social media is almost impossible, researchers are now concentrating on combating fake news on the Internet and social media to mitigate the enormous damage the spread of such news can cause to individuals, communities, and nations. To detect whether news is fake and stop it before it can spread, a reliable, rapid, and automated system using artificial intelligence should be applied. Hence, in this study, an Arabic fake-news detection system that uses machine-learning algorithms is proposed. An in-house Arabic dataset containing 206,080 tweets was collected using an API search on Twitter. The algorithm uses term frequency-inverse document frequency to extract features from the dataset and analysis of variance to select subsets from them. Nine machine-learning classifiers were used to train the model (naïve Bayes, K-nearest-neighbours, support vector machine, random forest (RF), J48, logistic regression, random committee (RC), J-Rip, and simple logistics). The experimental results indicated that the highest accuracy (97.3%) was obtained using the random forest and random committee, with training times of 4403s and 0.367s, respectively.

Keywords: *Cybersecurity, Artificial intelligence, Social Media, Fake News, Machine Learning, API Search, Twitter.*

1. INTRODUCTION

The Internet has played a fundamental role in many areas, including e-services, which have been adopted by many countries to provide services to their citizens [1], and e-government services, which address the needs of citizens, businesses, and the government sector [2].

Social media has become an essential part of our daily lives, serving as a source of information about events and real-time news (both local and global). People currently prefer online social media over traditional news and digital platforms such as television, radio, landline phones, and newspapers owing to their ease of use and widespread adoption worldwide. Online social media provide significant opportunities for business owners by facilitating online marketing, allowing them to sell their products and create instant connections with consumers at any time and from any country. However, social media has disadvantages, such as the spreading of harmful rumours in the form of fake news, information forgery, and falsified ratings and feedback. Additionally, social media can expose users to malware applications, which are major security threats [3] affecting the integrity, availability, and confidentiality of mobile systems. The cost of cybercrime worldwide was approximately $600 billion in 2018 [4], where Legal obstacles are frequently cited as one of the most important factors determining the efficiency of the global fight against cybercrime. Several researchers investigated the behaviour of malicious mobile applications allowing hackers to exploit synchronisation vulnerabilities when launching their attacks [5, 6] and the weaknesses of countermeasures such as anomaly- and signature-based detection [7-9]. According to a 2021 survey [10], media trust decreased by 8% worldwide between 2020 and 2021. Statistics indicate that an increasing number of people are losing faith in mainstream media every year. During the last three months of the 2016 U.S. presidential campaign, the number of interactions with fake-news stories increased from 3 to 8.7 million [11]. This statistic forced Facebook to prevent the spread of fake news on its platform. They reduced fake-news engagements from 200 million in 2016 to 70 million in 2018 by removing approximately 837 million spam posts and 583 million fake accounts during the first quarter of 2018 [12].

Fake-news publishers, who use online social media to spread false or misleading information, are among the most severe threats to such media. People who spread fake news may do so for political reasons, financial gain, advertising, or to harm the reputation of an individual. Rumours threaten democracy and freedom of expression because they can rapidly change public opinion and lead to distrust in governments and political conflicts [13]. For instance, fake news regarding measles led to an epidemic of the disease and the death of approximately 90% of the Amerindian population during the 15th century [14].

The rise of fake news and its significant negative impacts on individuals, communities, and governments have prompted researchers and governments to launch campaigns to combat it. Fake-news detection is complicated because of the large number of fake news stories that circulate globally and the complexity of languages and accents, making the examination of a single piece of news impossible. The ideal solution is to use artificial-intelligence (AI) techniques to learn from user behaviour and detect fake news on social media platforms.

Over the past decade, fake-news detection and classification have attracted the attention of many researchers. Most [15-17] have focused on classifying news in English and a few European languages [18, 19]. Few studies have focused on Arabic content [20-22], even though Arabic is the official language of 25 countries and the mother tongue of over 466 million people and with over 30 accents. Compared with English, Arabic is more complicated and has complex morphologies, making it difficult to learn. User-generated content on the Internet has additional complexity because most people write in their native tongue rather than in the official language. In this study, we focused on fake-news classification for breaking news owing to its tendency to spread rapidly. We also focused on Arabic fake-news classification using machine-learning methods.

The remainder of this paper is organised as follows. In Section 2, we present several related research efforts, and in Section 3, we review the methodology and outline the various machine-learning algorithms used to build the classification model. The classification results are presented in Section 4. In Section 5, we discuss the main contributions of the study. Finally, concluding remarks are presented in Section 6.

2. RELATED WORK

This section provides an overview of existing fake-news detection techniques. Examples of fake news in Arabic, English, and other languages are presented to illustrate the differences and similarities among the methods and their effects on the performance.

2.1 Fake News in Arabic

With the increasing and widespread dissemination of fake news in Arabic countries, Arabic fake-news classification has attracted the attention of researchers. In [20], a novel Arabic corpus for analysing fake-news tasks was introduced. The researchers focused on fake news concerning the deaths of three popular Arab celebrities: the comedian Adel Imam, President Bouteflika, and the dancer Fifi Abdou. They collected 4079 related stories, divided into three categories according to the celebrity's name using the YouTube API. They improved the data quality by removing noise, such as particular characters, URL links, non-Arabic words, and duplicate comments. They then used three machine-learning classifiers: support vector machine (SVM), decision tree (DT), and multinomial naïve Bayes (MNB1). Next, they split the data into a training set (70%) and a test set (30%). In the tests, the highest accuracy (95.56%) was achieved using the DT classifier.

The authors of [23] and [21] used many feature-extraction techniques, for example, term frequency-inverse document frequency (TF-IDF) and word embedding, to extract valuable features from datasets. In [22], a ClaimRank model was developed for detecting check-worthy claims by using seven English datasets and translating two of them into

Arabic. Then, features were extracted using techniques, such as TF-IDF, weighted bag of words, part-of-speech tags, sentiment scores, and sentence length (in tokens). The authors also added a language detector for Arabic adaptation, after which they used a neural network (NN) with two hidden layers to train their model.

Similarly, in [24], the TF-IDF method was used to identify whether new tweets were fake news. The model employed was based on the cosine similarity technique used to measure the credibility of tweets on Twitter. Approximately 700 rumour tweets on sensitive topics, such as politics, health, and crises, were collected, and news from official Twitter accounts, such as @AJArabic and @cnnarabic. The data were pre-processed by removing stop words and deleting redundant data, and then features were extracted using the TF-IDF technique. The system detected 67% of the rumours and 80% of the news stories.

The authors of [21] proposed a model consisting of several stages to detect misinformation regarding COVID-19 in social media using machine-learning and deep-learning techniques. First, they constructed a large Arabic dataset related to COVID-19 by collecting more than 4,514,136 tweets using the Tweepy Python library and the Twitter streaming API. They then used pre-processing methods to clean the data and remove unwanted parts, such as non-Arabic words, special characters, URLs, and punctuation marks. They also used the TextBlob Python library to conduct a text correction, normalise the Arabic text, remove repeated characters and stop words, and apply word stemming. They then extracted the features from the data using two feature-extraction techniques: TF-IDF and word embedding. Finally, they applied five machine-learning classifiers, i.e. random forest (RF), extreme gradient boosting (XGB), naive Bayes (NB), stochastic gradient descent (SGD), and SVM, and three deep-learning classifiers, i.e. a convolutional neural network (CNN), recurrent neural network (RNN), and convolutional recurrent neural network (CRNN). The SVM classifier achieved the highest accuracy (87.8%).

In [25], a system comprising four main modules for detecting fake Arabic news on Twitter was proposed. The first module applies feature extraction and content parsing; the second, content verification; the third, a polarity evaluation of user comments; and the fourth, credibility classification. Approximately 800 Arabic news items were collected and labelled manually, and then the user-based, content-based (CB), and sentiment features were extracted from the data. The authors processed the collected data using three machine-learning classifiers: DT, SVM, and NB. Their system achieved an accuracy of 89.9% using a DT classifier.

2.2 Fake News in English

The authors of [16] proposed a simple model for detecting fake news using the NB classifier and used a dataset collected by BuzzFeed containing 2282 posts (1145 posts from mainstream pages, 471 from left-wing pages, and 666 from right-wing pages). Their system achieved an accuracy of 74%, which is reasonable considering its simplicity. The results can be improved by using a larger dataset and applying pre-processing techniques, such as removing stopping words.

One main challenge in building a detection system is finding a suitable dataset. Some datasets are private or confidential. Therefore, researchers have preferred to create custom datasets. In [15] and [17], manually collected datasets were used for fake-news detection. In [17], a weakly supervised model was developed by collecting a large-scale training dataset using the Twitter API. The dataset contained thousands of tweets labelled automatically during the collection phase. The authors used five feature-extraction techniques, based on user-level, tweet-level, text, topic, and sentiment features. The system was evaluated using two settings: cross-validation and validation against the gold standard. Although the dataset was not cleaned and was inaccurate, the system detected fake news with an F1-score of 90%. In [15], machine learning techniques were used to build a fake-news detection model. The authors collected 948,373 messages using a Twitter API and cleaned the data by removing the replicated data after normalising them. They used three popular classifiers: NB, an NN, and an SVM. The model achieved an accuracy of 99.08% using an NN and an SVM.

Previously reported fake-news detection models are based on linguistic features [24, 25]. The authors of [26] employed CB features and machine-learning algorithms to build a fake-news detection model using several linguistic feature sets. They used two feature-selection techniques to select the valuable features, i.e. mutual information and mRMR, and their experimental results indicated that the mutual information yielded better results than mRMR. They also introduced UNBiased—a text corpus dataset containing 1400 texts labelled by experts as fake and 2004 labelled as real—and used simple classifiers (NB, SVM, DT, k-nearest neighbours (KNN)) and ensemble classifiers (AdaBoost and Bagging). Their system achieved a high accuracy of 95% by using ensemble algorithms and an SVM and a linguistic feature set with word embeddings. In another study [27], a linguistic model was proposed for extracting grammatical, syntactic, sentimental, and readability features from the news. Two datasets were used to evaluate the model: 1) BuzzFeed Political News Data, containing 48 fake and 53 real news items, and 2) random political news data, containing 75 fake and 75 real news items. NN and LSTM deep classifiers trained the model and achieved an accuracy of 86%.

In another study, the TF-IDF feature-extraction method was used to build a detection model [28]. The authors proposed a system comprising three levels: 1) pre-processing techniques, such as removing stop words, lowercase characters, punctuation marks, numbers, special characters, and white spaces; 2) features extracted from the dataset; and 3) as multi-layer perceptron (MLP) with three layers, i.e. an input layer, a hidden layer, and output layer, using feedforward and backpropagation as classifiers. They used the *FNs articles.CSV* file dataset collected from the kaggel.com website, which contains 10,423 for real news and 10,432 for fake news. Their system achieved an accuracy of 95.47%.

2.3 Other Languages

Fake-news classification is a global issue, and the fight against it has attracted the attention of governments and researchers worldwide. Although most researchers have focused on English, many have proposed systems for detecting fake news in foreign languages. The authors of [29] proposed the SpotFake multi-model for fake Chinese news detection, which exploits the visual and textual features of the articles. They used a pre-trained bidirectional encoder representation from transformers (BERT) model to extract features from news text (textual features) and VGG19 to extract features from image data (visual features). Additionally, they used two public datasets: the Twitter MediaEval dataset, which contains 17,000 tweets (9000 fake tweets, 6000 real news tweets, and 2000 test news tweets), and the Weibo dataset, which is a collection of real news from authoritative news sources in China, including the Weibo microblogging website and Xinhua News Agency. The fake news was collected from Weibo from May 2012 to June 2016. The authors evaluated their approach using the previously developed EANN [30] and MVAE [31] models, and their system achieved accuracies of 89.23% using the Weibo dataset and 77% using the Twitter dataset.

In other studies [17, 18], the fake-news classification problem was solved for Italian news articles. The authors of [18] proposed a simple approach based on the multi-layer representation of Twitter networks, where each layer represents one type of interaction, such as a tweet, mention, or retweet. They used two large datasets: 1) a U.S. dataset containing 2,039,098 mainstream Twitter and 1,667,807 disinformation interactions associated with the most trusted sources from a dozen U.S. mainstream news websites collected from 25 February to 18 March 2019 using the Streaming API and 2) and an Italian dataset containing 27,055 mainstream and 44,932 disinformation interactions collected from 19 April to 5 May 2019, also using the Streaming API. They applied a logistic regression (LR) classifier with an L2 penalty, for which their system obtained an area under the receiver operating characteristic (AUROC) of up to 94%. The authors of [19] combined social and news contents features to develop an HC-CB-3 detection system that outperformed previously proposed methods by up to 4.8%. They used FacebookData [32], which contains 15,500 posts from 32 pages; datasets collected from FakeNewsNet containing 240 news items labelled by the PolitiFact fact-checking site; and a dataset containing 182 news items labelled by BuzzFeed. They used LR with CB and harmonic crowdsourcing (HC). They implemented a trained chatbot to classify the news as fake or real and then tested it on real-world data and achieved an accuracy of 81.7%. The HC-CB-3 system achieved an accuracy of 99.1% using the FacebookData dataset.

Other researchers proposed systems for detecting fake news written in floating language types; for example, in [33] a detection system based on morphological analysis for detecting fake articles written in the Slovak language was presented. The authors built a dataset by collecting 160 articles from fake-news publishers and then used a morphological analysis to pre-process the data and improve the results. The authors used a DT classifier and achieved the highest accuracy (75%) with a maximum tree depth of 9.

Table 1 Fake-news detection methods using ML for different languages

Reference	Language	Dataset	Method	Classifier	Feature Extraction	Accuracy
[20]	Arabic	Collected 4,079 stories	Introduced a novel Arabic corpus for analysed the fake-news tasks that concerned the death of three Arab celebrities	SVM, DT, MNB	---	95.56%
[21]	Arabic	Seven English datasets, and two Arabic datasets	Proposed a ClaimRank model to detect the check-worthy claims	Neural Network	TF-IDF, weighted bag of words	---
[23]	Arabic	Collected 4,514,136 million tweets	Proposed a model that consists of several stages to detect the misinformation about the covid-19 in social media based on machine learning and deep learning techniques	RF, XGB, NB, SGD, SVM, CNN, RNN, CRNN	TF-IDF and word embeddings	87.8%
[22]	Arabic	700 rumours tweets	Proposed model that based on the cosine similarity technique to measure credibility of tweets in the twitter	---	TF-IDF	80%
[23]	Arabic	Collected 800 Arabic news	proposed a system that consist of four main modules to detect Arabic fake news on the twitter	DT, SVM, NB,	Twitter API	89.9%
[15]	English	2282 posts	Proposed a simple model to detect fake news using the Naïve Bayes classifier	Naïve Bayes	---	74%
[16]	English	Collected thousand tweets	Proposed a weakly supervised model that collected a large-scale training dataset contains a thousand tweets that labelled automatically during the collection	NB, DT, SVM, NN	user-level features, tweet-level features, text features, topic features, and sentiment features	90%
[14]	English	Collected 948373 messages	Used the machine learning techniques to build a fake news detection model,	NB, NN and SVM	---	99.08%
[24]	English	UNBiased contains 1400 fake and 2004 real	Used content-based features and Machine Learning algorithms to build a fake news detection model using several linguistic feature sets	NB, SVM, DT, KNN, Ada Boost and Bagging	linguistic feature sets, word embeddings	95%
[25]	English	1-Buzzfeed contains 48 fake and 53 reals. 2- Random Political News Data contains 75 fake and 75 reals	Proposed a linguistic model to extract grammatical, syntactic, sentimental and readability features from the news.	Neural Network and LSTM	linguistic feature sets	86%
[26]	English	FNs articles.CSV file contains 20,800 records	Used the TF-IDF and MLP with three layers to build a detection model	MLP	TF-IDS	95.47%
[27]	Chinese	Twitter MediaEval and Weibo dataset	Proposed the SpotFake multi-model for Chinese fake news detection that exploits the visual and textual features from the articles	BEERT and VGG19	BEERT and VGG19	89.23%
[17]	Italian	US dataset and Italian dataset	Proposed a simple approach based on the multi-layer representation of twitter networks	Logistic Regression	---	94%.

[18]	Italian	FacebookData, PolitiFact, BuzzFeed	Combined social content and news content features to propose a HC-CB-3 detection system	Logistic Regression	---	99.1%
[31]	Slovak	Collected 160 articles	Presented a detection system based on morphological analysis to detect the fake articles that written in Slovak language	Decision Tree	---	75%

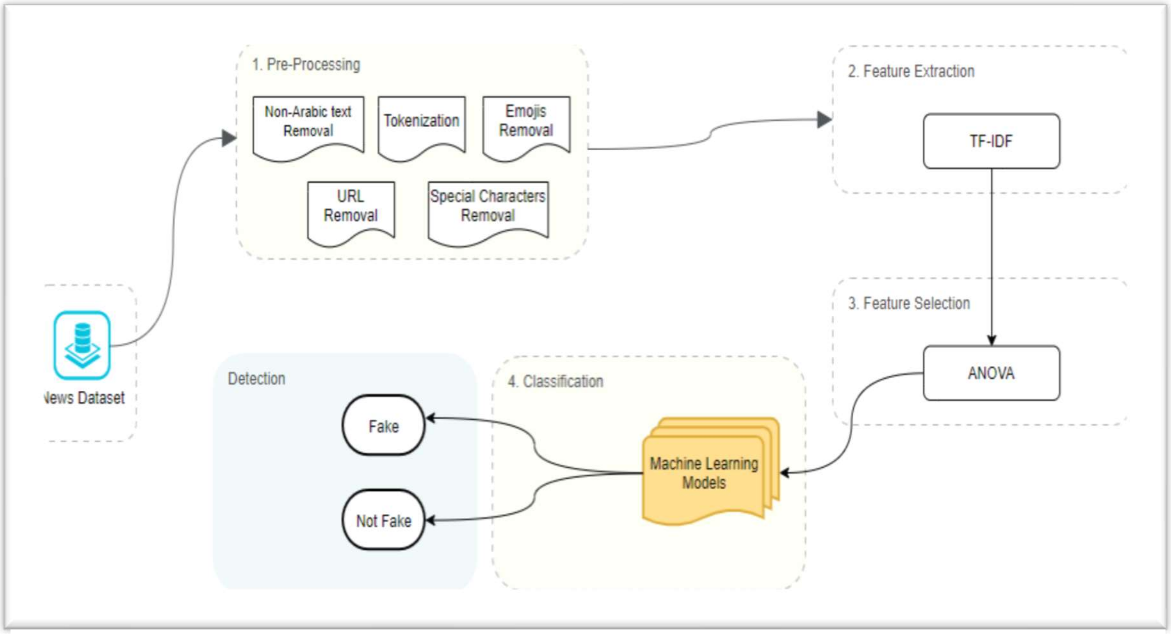


Figure 1: Overview of the fake-news classification methodology

Other researchers proposed systems for detecting fake news written in floating language types; for example, in [33] a detection system based on morphological analysis for detecting fake articles written in the Slovak language was presented. The authors built a dataset by collecting 160 articles from fake-news publishers and then used a morphological analysis to pre-process the data and improve the results. The authors used a DT classifier and achieved the highest accuracy (75%) with a maximum tree depth of 9.

3. METHODOLOGY

Figure1 illustrates the elements of the classification methodology used in this study.

1. Dataset: The first and most crucial step in the classification methodology is to find a suitable dataset; in this study, we collected 206,080 tweets.
2. Pre-processing: In this step, we pre-processed the data, including non-Arabic text removal, tokenisation, emoji removal, special-character removal, and URL removal.
3. Feature Extraction: We used the TF-IDF algorithm to generate informative values from the dataset.
4. Feature Selection: Analysis of variance (ANOVA) was used to select sets of 50, 75, 100, 150, and 200 features.
5. Classifier: The classification system was built using nine machine-learning algorithms (NB, KNN, SVM, RF, J48, LR, random committee (RC), J-Rip, and simple logistics).
6. Training Algorithm: We trained the model by classifying the news as rumours or non-rumours using SGD.
7. Performance Evaluation: The proposed model was tested via 10-fold cross-validation.

3.1 Dataset

We created a custom news dataset to evaluate the performance of our classifiers and test the effectiveness of the proposed system by collecting 206,080 tweets related to fake and real topics using the Twitter API. The fake topics were selected in accordance with the Anti-Rumour Authority, which was formed in 2012 to combat the spreading of fake news on social media [4]. Table 2 presents a sample of the collected tweets, and Table 3 presents the dataset statistics.

Table 2 Sample of the collected dataset

Table 3 Dataset statistics

Fake	**46796**
Clean	**159284**
Total	**206080**

3.2 Pre-Processing

Pre-processing is an important phase of any machine-learning algorithm [34]; it is necessary for removing all noise from the data before extracting the features to improve the system performance. We used the following pre-processing steps in this study:

- Tokenisation: Each tweet was split into a sequence of words or tokens according to white spaces.
- Non-Arabic Text Removal: Each token was examined to ensure that all non-Arabic text was removed.
- Emoji Removal:Emojis were removed from all tweets to reduce the amount of noise.
- Special-Character Removal: All special characters (e.g. @, *, %, ^, &) were removed.
- URL Removal: The URL links were removed.

3.3 Feature Extraction

In this study, the TF-IDF [35] feature-extraction technique was used to evaluate the importance of the words that appeared in the documents by counting them using (1). The TF-IDF is a product of the term frequency (TF) and the inverse document frequency (IDF). A high score is obtained when the term has a high frequency in a document having a low frequency in the corpus.

$$tf-idf(w)=\frac{n*count(w)}{\ln\left(\frac{n}{m+1}\right)*\sum_{j}^{m}count(w^{t_j})} \qquad (1)$$

Here, *w* represents the word, $count(w^{t_j})$ represents the number of times *w* appears in the corpus, and *m* represents the number of samples that contain w in the corpus.

3.4 Feature Selection

All the extracted features cannot be used for the training phase, as the classifier may be confused if some of the features are noisy and redundant [36], resulting in a slow training process. ANOVA was used to select the relevant and highest-scoring features to train the model by measuring the similarity between pertinent features and reducing the scale of the feature vectors between fake and non-fake news. ANOVA reduced the features to sets of 50, 75, 100, 150, and 200, which are the sizes most often used in the literature, via the following equation [4]:

$$\sum_{i=1}^{k}\frac{n_i(\bar{Y}_i-\bar{Y})^2}{(K-1)}. \qquad (2)$$

News Type	News
Fake	بعد القصبي ناصر الممثل تعالى الله رحمة الى انتقل يرحمه الله القصيم طريق على حادث الى تعرضه القصبي ناصر وفاة
Fake	هبوط الجويه الاحوال سوء بسبب السعوديه | عاجل السريع الخط على جدة و مكة بين قليل قبل طائرة
Not Fake	حديدي خط : الإماراتية المواصلات هيئة عاجل م2021 ديسمبر في السعودية بـ الإمارات يربط
Not Fake	تاريخها في للنفط حقل أكبر اكتشاف تعلن البحرين

3.5 Machine-Learning Classifiers

Machine learning is a subfield of AI in which computers employ statistical techniques to learn [37]. Two types of learning are commonly used: supervised and unsupervised. The classification of fake news falls under supervised learning [38], which involves mapping an input to an output according to labels.

In this study, we used nine machine-learning classifiers to classify the news as fake or not fake: SVM, KNN, NB, RF, RC, J48, J-Rip, LR, and simple logistics.

3.5.1 Naïve Bayes

Naïve Bayes (NB) is a probabilistic classifier with a collection of algorithms, based on the Bayes theorem, and states that every pair of features can be classified independently of the others [39]. There are three types of NB classifiers based on these features: Gaussian, multinomial, and Bernoulli. NB is a classification algorithm that is widely used to solve classification problems. Its applications include spam filtering, text analysis, and recommendation systems.

NB can be measured using the following equation:

$$P(X|C_i)=\frac{1}{\sqrt{2\pi\sigma}}e^{\frac{(X-\mu)^2}{2\sigma^2}}, \qquad (3)$$

where μ represents the mean, and σ^2 represents the variance.

3.5.2 K-Nearest-Neighbours

KNN is a basic supervised learning algorithm and is used to solve classification problems by measuring the similarity and distance between the input and testing data using a distance function (e.g. the Euclidean distance equation) and classifying the data according to the nearest neighbour point. It requires a long time to predict the distance between all data points owing to its simplicity and incorrectly classifies data located near the boundaries[40]. The following equation is used to calculate the Euclidean distance:

$$d(a,b) = d(b,a) = \sqrt{\sum_{i=1}^{n}(b_i - a_i)^2}. \qquad (4)$$

3.5.3 Support Vector Machine

An SVM is a robust supervised algorithm that solves classification, regression, and outlier-detection problems. It uses the number of features (N) to find the hyperplane in the N-dimensional space and classify unlabelled data points [41]. A linear SVM divides data points into two classes according to a straight line created between the classes; it draws an infinite number of lines to find the one that is farthest from the closest data points. The main issue with SVMs is that they do not provide a direct estimation, necessitating five-fold cross-validation at minimum, which is computationally expensive. An optimising SVM can be calculated using the following equation [42]:

$$\min_{w \in \mathbb{R}^d} C \sum_{i}^{n} \max\left(0, 1 - y_i f(x_i)\right) + \|w\|^2. \qquad (5)$$

3.5.4 Random-Forest

RF is a meta-estimator composed of numerous DTs, and it prevents overfitting and improves the accuracy through averaging. It is based on the wisdom of crowds; i.e. each DT predicts a class, and then the decisions are merged, and the class with the most votes is selected as the model prediction, resulting in ensemble learning. It uses the entropy formula or the Gini equation [43] to determine the nodes on a DT branch. An RF averages the prediction using the following equation [44]:

$$\hat{y} = \sum_{i=1}^{n}\left(\frac{1}{m}\sum_{j=1}^{m} w_j(x_i, \acute{x})\right), \qquad (6)$$

where $(\hat{y})$ represents the prediction, m represents the set of trees, and w_j represents the individual weight function.

3.5.5 J48

J48 is one of the most effective supervised learning algorithms for analysing categorical and continuous data using the information entropy formula [45]. J48 determines how nodes in the DT should be branched according to the outcome probability, used to generate the DT tool. The main issue with the J48 classifier is that it usually requires a large space complexity because it relies on the depth between the root and leaves.

$$E(y) = \sum_{i=1}^{\#of\ classes} -p_i \log_2(p_i). \qquad (7)$$

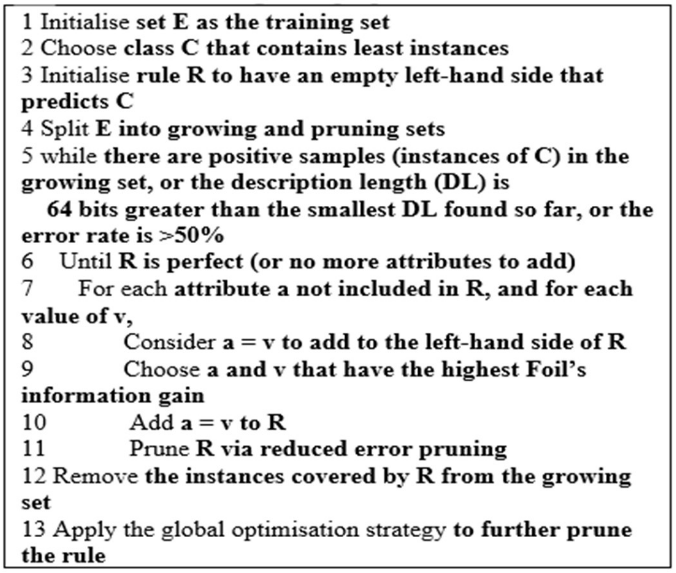
Here, p_i represents the probability, and the summation represents the sum of the possible values.

3.5.6 Logistic Regression

LR, also known as the sigmoid function, is a foundational supervised learning algorithm used to solve binary classification problems. It reduces the continuous input values to the range of (0,1), making it helpful in dealing with probabilities, such as predicting whether the news is fake (0) or real (1). LR involves introducing a nonlinear form by learning a linear relationship from a labelled dataset and categorising it into its classes. It is implemented using the following equation [46]:

$$P(Y|X) = \frac{1}{1+e^{-f(x)}}, \qquad (8)$$

where *f*(*x*) is a function containing the features *x*.



```
Initialise set E as the training set
Choose class C that contains least instances
Initialise rule R to have an empty left-hand side that predicts C
Split E into growing and pruning sets
while there are positive samples (instances of C) in the growing set, or the description length (DL) is
    64 bits greater than the smallest DL found so far, or the error rate is >50%
  Until R is perfect (or no more attributes to add)
    For each attribute a not included in R, and for each value of v,
      Consider a = v to add to the left-hand side of R
      Choose a and v that have the highest Foil's information gain
      Add a = v to R
      Prune R via reduced error pruning
Remove the instances covered by R from the growing set
Apply the global optimisation strategy to further prune the rule
```

Figure. 2 Procedures of J-Rip algorithm

3.5.7 Random Committee

RC is a supervised machine-learning algorithm used in a meta-classifier classification [47]. It is an ensemble-based classifier because it is implemented by combining different random trees with varying numbers of seeds, all of which use the same training dataset. The final prediction is generated by averaging the base classifiers' probability estimations.

3.5.8 J-Rip

Cohen [48] created J-Rip, which is an optimised version of IREP. J-Rip is based on rule association and applies the repeated incremental pruning to produce the error reduction (RIPPER) concept by deriving a set of rules from the training set. The main advantage of J-Rip is that it works well with noisy datasets with imbalanced classes. The procedure of J-rip is discussed in figure 2 [49].

4. RESULTS

The experimental results are presented in this section.

4.1 Evaluation Metrics

The evaluation metrics presented below were used to assess the effectiveness of the proposed approach.

1. **Precision (positive predictive value)**: This is the percentage of relevant results, which quantifies the number of positive prediction classes that truly belong to the positive class [50]. It is calculated as follows [51]:

$$Precision = \frac{TP}{TP+FP}, \quad (9)$$

where *TP* represents a true positive, and *FP* represents a false positive.

2. **Recall (Sensitivity):** This is the percentage of the total results that are relevant and classified correctly using the proposed model [52]. It is calculated as follows [53]:

$$Recall = \frac{TP}{TP+FN}, \quad (10)$$

where *FN* represents a false negative.

3. **F-measure (F1-Score):** This is a function of the recall and precision [54] and is used to measure the system accuracy for a dataset. It is calculated using the following equation [55]:

$$F - measure = 2 * \frac{Precision*Recall}{Precision+R}. \quad (11)$$

4. **Area under the curve (AUC):** This is a performance measurement method used to evaluate the ability of a classifier to discriminate between classes.
5. **Receiver operating characteristic (ROC) curve:** This plot represents the classification model performance at all classification thresholds**.**
6. **Learning curve**: This plot presents the model performance by diagnosing an underfit, overfit, or well-fit model for the validation and training datasets.

4.2 Evaluation Metrics

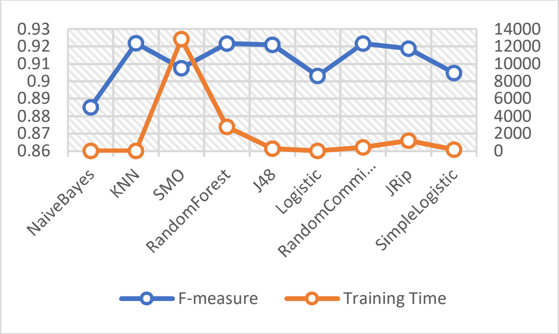
In our experiments, we used the ANOVA feature-selection algorithm with nine different classifiers: NB, KNN, SVM, RF, J48, LR, RC, J-Rip, and simple logistics. The F-measure and elapsed training time were used to evaluate the classification model. ANOVA was performed to reduce the features to sets of 50, 75, 100, 150, and 200.

The highest accuracy (97.3%) was achieved by employing an RF and RC with a 200-feature set, as shown in Figure 6. The training times required by the RF and RC were 4403 and 525 s, respectively. Thus, the RF took longer to obtain the classification results, and the RC exhibited better performance in terms of the accuracy and time complexity. The RC classifier achieved the highest accuracy. Figure3 shows the performance of the proposed model in terms of the precision, recall, and F-measure when different numbers of features were selected. The highest precision (97.38%), recall (97.35%), and F1-score (97.3%) were achieved using the 200-feature set, as shown in Figures 3(a)–3(c), respectively.

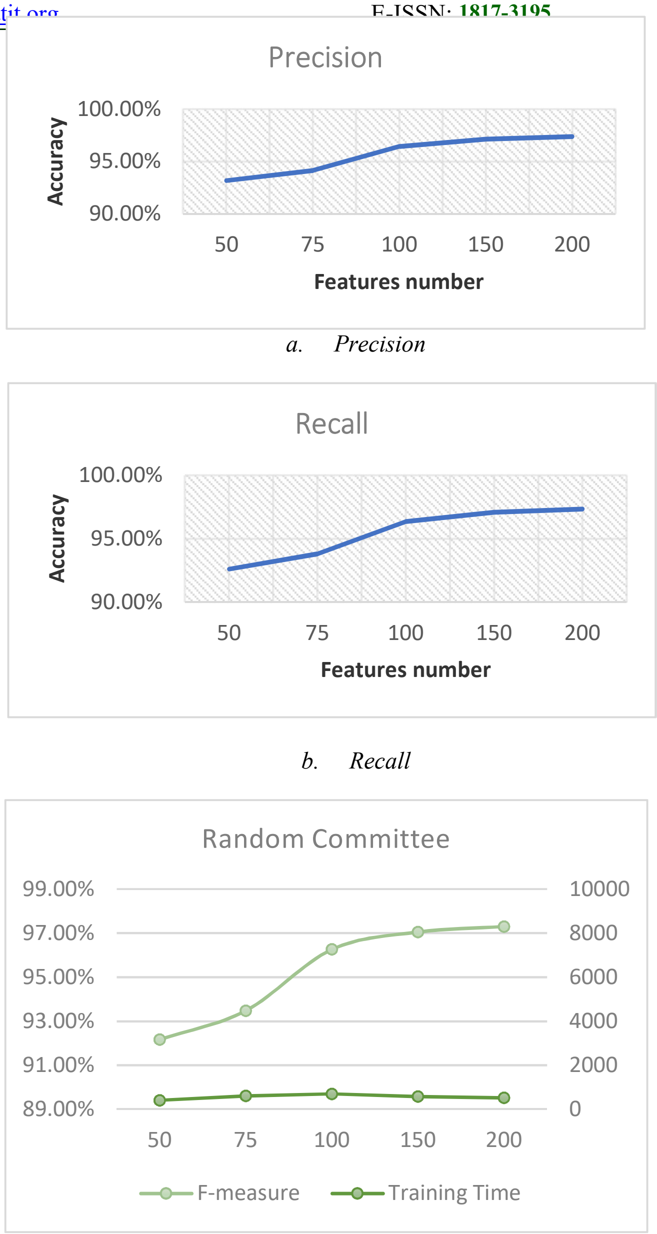
Figure 4 shows the model performance when different feature sets are selected by using the ANOVA, which improves the accuracy of our model. The sub-figures indicate that the number of selected features selected affecting the system performance; it is directly proportional to the accuracy and inversely proportional to the training speed. The classification performance improves as the number of features selected increases. As shown in Figure 4(a), when a 50-feature set is selected, the highest accuracy (92.1%) is obtained using the KNN, RF, and RC classifiers, with complexity training times of 0.096, 2755, and 408 s, respectively. Figure 4(b) shows the performance of the model when a 75-feature set is selected; the KNN and RF classifiers achieve the highest accuracy (93.5%). KNN is the most efficient classifier when 100 features are selected, owing to its high accuracy and low training time, as shown in Figure 4(c). Figures 4(d) and 4(e) present the model performance when 150- and 200-feature sets, respectively, are selected. As shown, the highest accuracy is achieved in these cases

5. DISCUSSION

The contributions of this study are as follows. First, we compared the performance of different classifiers for detecting fake news. The RF and RC classifiers obtained the same F1-score (97.3%), whereas the RF surpassed the RC in terms of the AUC score. The second contribution was the comparison of various input feature sets selected using the ANOVA feature-selection technique



(a) Selecting 50-feature set

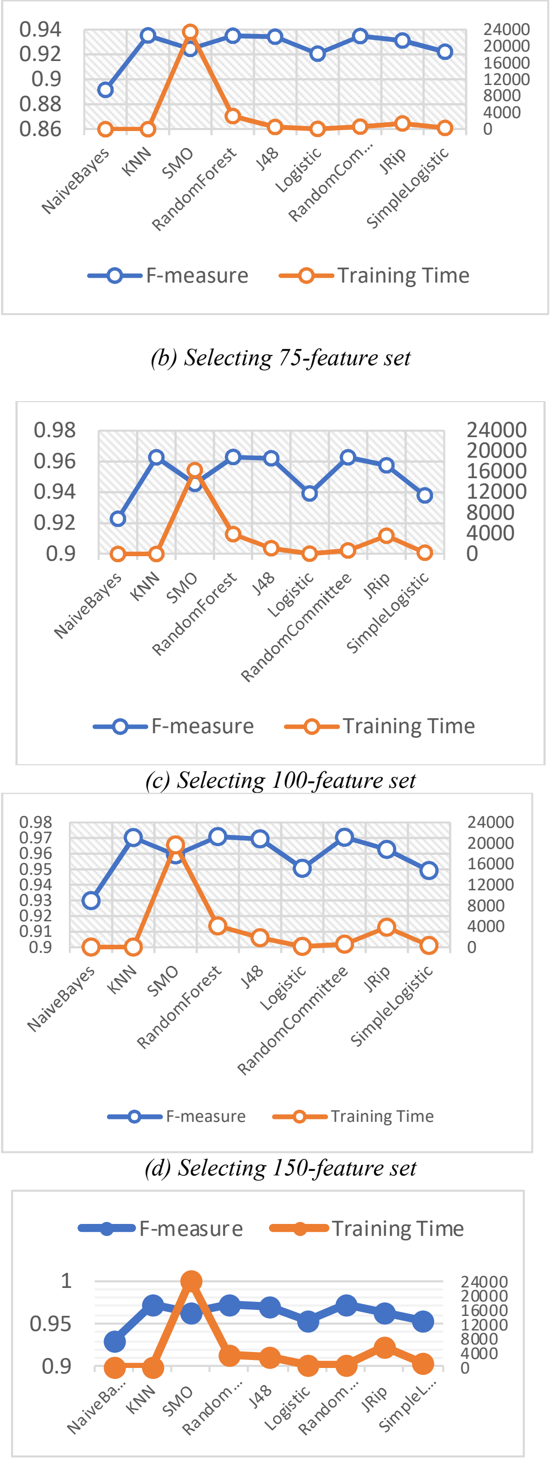


a. Precision

b. Recall

c. F-measure and training time

Figure 3. A Set Of Three Sub-Figures Showing How The Evaluation Metrics Improve As The Number Of Selected Features Increase When Training The Model Using An RC Classifier With A 10-Fold Cross-Validation



(b) Selecting 75-feature set

(c) Selecting 100-feature set

(d) Selecting 150-feature set

(e) Selecting 200-feature set

Figure 5. A set of five sub-figures showing the classification accuracy and required training time when selecting 50-, 75-, 100-, 150-, and 200-feature set

The ROC curve is useful for representing the relationship between the true positive rate and the false positive rate. The RF achieved the best ROC (0.99) in our experiments when the curve had the shortest distance to the upper-left corner, as shown in Figure 5.

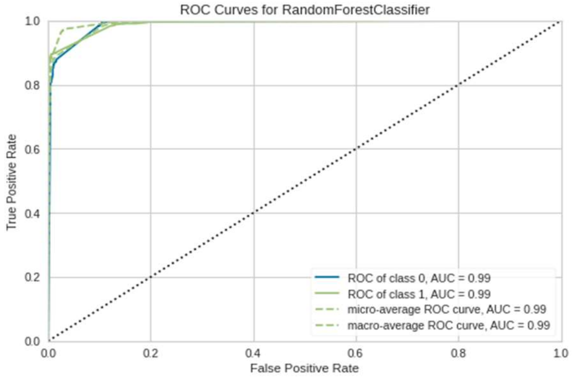


Figure 6. ROC curves of the RF classifier

Figure 6 presents the learning curves for the RF classifier. The cross-validation and training curves converge at a high score, indicating that the model does not suffer from overfitting or underfitting issues and that increasing the training size increases the accuracy of the model.

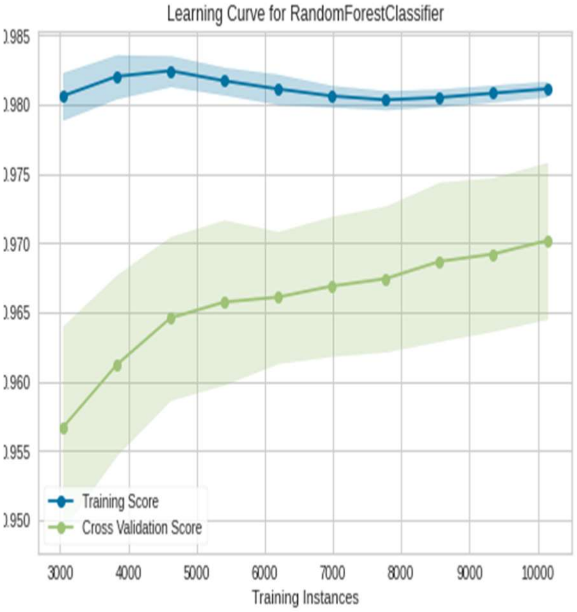


Figure 6. Learning curve for the RF classifier

As shown in Figure 7, the confusion matrix for the RF-classifier results is examined. A total of 802 instances is correctly classified as true positives, and 122 positive instances are incorrectly classified as negative. Furthermore, 3072 instances are correctly classified as negative, and 27 negative instances are incorrectly classified as positive.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

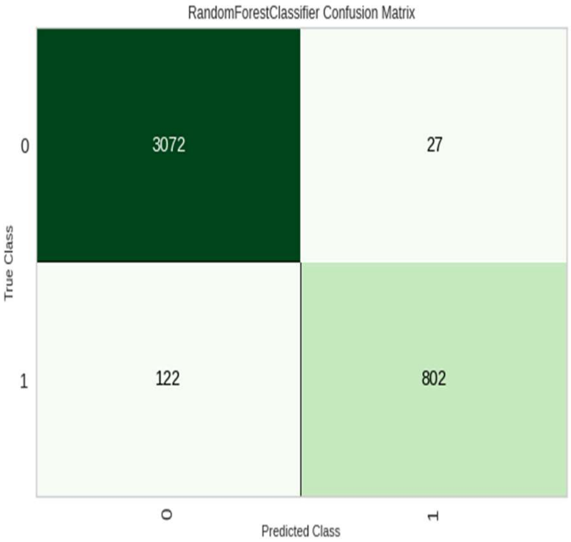


Figure 7. RF-classifier confusion matrix

'Feature importance' refers to methods that assign a score to the input features according to their ability to predict the variables. This concept is helpful in regression and classification problems; it leads to a thorough understanding of the problem under consideration and enables the development of more accurate and efficient classifiers. The highest accuracy was obtained using a 200-feature set, and the importance of the top 10 features chosen from this set is presented in Figure 8.

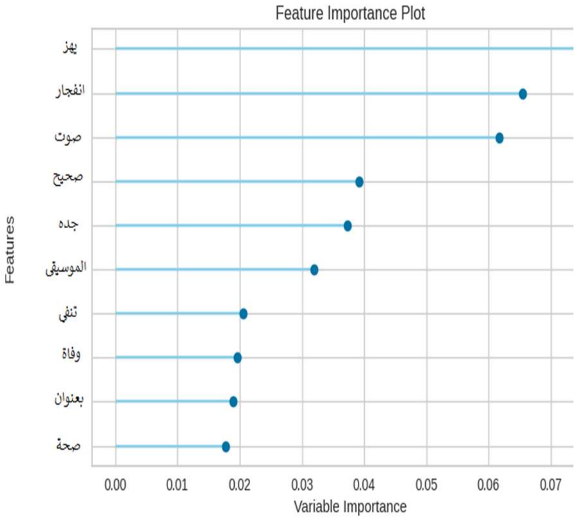


Figure 8. Importance of top-10 features selected from 200-feature set

6. CONCLUSION

Fake news is defined as misleading or deceptive information. The main concern with fake news is that it spreads quickly—particularly on social-media platforms such as Twitter and Facebook [56]. In this paper, we proposed a novel technique for determining whether Arabic news is fake. The proposed technique is based on a combination of text-mining methods and has four major phases: data collection, pre-processing, processing, and system evaluation. We collected 206,080 tweets related to fake and non-fake events using the Twitter Search API and then used pre-processing techniques such as non-Arabic text removal, tokenisation, and emoji, special-character, and URL removal to eliminate unwanted noise. We then employed the TF-IDF algorithm to select a total of 64,447 features. The ANOVA method was used to determine the top 200, 150, 100, 75, and 50 features, and we then used nine classifiers to train the proposed model. For the evaluation, a 10-fold cross-validation method was applied. The highest accuracy (97.3%) was achieved using a 200-feature set with the RF and RC classifiers.

REFERENCES:

[1] Alhyari, S., et al., *Six Sigma approach to improve quality in e-services: An empirical study in Jordan.* International Journal of Electronic Government Research (IJEGR), 2012. **8**(2): p. 57-74.

[2] Alhyari, S., et al., *Performance evaluation of e-government services using balanced scorecard: An empirical study in Jordan.* Benchmarking: an international journal, 2013.

[3] Alazab, M., et al., *Cybercrime: the case of obfuscated malware*, in *Global security, safety and sustainability & e-Democracy*. 2011, Springer, Berlin, Heidelberg. p. 204-211.

[4] Alazab, M., et al., *Intelligent mobile malware detection using permission requests and API calls.* Future Generation Computer Systems, 2020. **107**: p. 509-521.

[5] Alazab, M., et al. *Analysis of malicious and benign android applications*. in *2012 32nd International Conference on Distributed Computing Systems Workshops*. 2012. IEEE.

[6] Alazab, M., A. Alazab, and L. Batten. *Smartphone malware based on synchronisation vulnerabilities*. in *ICITA 2011: Proceedings of the 7th International Conference on Information Technology and Applications*. 2011. ICITA.

[7] Alazab, M. and L.M. Batten, *Survey in smartphone malware analysis techniques.* New threats and countermeasures in digital crime and cyber terrorism, 2015: p. 105-130.

[8] Alazab, M., et al., *Zero-day malware detection based on supervised learning algorithms of API call signatures.* 2010.

[9] Alazab, M., et al., *Information security governance: the art of detecting hidden malware*, in *IT security governance innovations: theory and research*. 2013, IGI Global. p. 293-315.

[10] Barometer, E.T., *Edelman.* Retrieved October, 2020. **7**: p. 2020.

[11] 11. Yousefi-Azar, M., et al., *Malytics: a malware detection scheme.* 2018. **6**: p. 49418-49431.

[12] Allcott, H., M. Gentzkow, and C. Yu, *Trends in the diffusion of misinformation on social media.* Research & Politics, 2019. **6**(2): p. 2053168019848554.

[13] Asghar, M.Z., et al., *Exploring deep neural networks for rumor detection.* J Ambient Intell Human Comput, 2021. **12**(4): p. 4315-4333.

[14] Tangvatcharapong, M., *The Impact of Fake News: Evidence from the Anti-Vaccination Movement in the US.* 2019.

[15] Aphiwongsophon, S. and P. Chongstitvatana. *Detecting fake news with machine learning method*. in *2018 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*. 2018. IEEE.

[16] Granik, M. and V. Mesyura. *Fake news detection using naive Bayes classifier*. in *2017 IEEE first Ukraine conference on electrical and computer engineering (UKRCON)*. 2017. IEEE.

[17] Helmstetter, S. and H. Paulheim. *Weakly supervised learning for fake news detection on Twitter*. in *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. 2018. IEEE.

[18] Pierri, F., C. Piccardi, and S. Ceri, *A multi-layer approach to disinformation detection in US and Italian news spreading on Twitter.* EPJ Data Science, 2020. **9**(1): p. 35.

[19] Della Vedova, M.L., et al. *Automatic online fake news detection combining content and social signals*. in *2018 22nd Conference of Open Innovations Association (FRUCT)*. 2018. IEEE.

[20] Alkhair, M., et al. *An arabic corpus of fake news: Collection, analysis and classification*. in *International Conference on Arabic Language Processing*. 2019. Springer.

[21] Alqurashi, S., et al., *Eating garlic prevents covid-19 infection: Detecting misinformation on the arabic content of twitter.* arXiv preprint arXiv:2101.05626, 2021.

[22] Jaradat, I., et al., *ClaimRank: Detecting check-worthy claims in Arabic and English.* arXiv preprint arXiv:1804.07587, 2018.

[23] Raff, E., et al., *Kilograms: Very large n-grams for malware classification.* 2019.

[24] Floos, A.Y.M., *Arabic rumours identification by measuring the credibility of arabic tweet content*, in *Media Controversy: Breakthroughs in Research and Practice*. 2020, IGI Global. p. 236-248.

[25] Sabbeh, S.F. and S.Y. BAATWAH, *ARABIC NEWS CREDIBILITY ON TWITTER: AN ENHANCED MODEL USING HYBRID FEATURES.* journal of theoretical & applied information technology, 2018. **96**(8).

[26] Gravanis, G., et al., *Behind the cues: A benchmarking study for fake news detection.* Expert Systems with Applications, 2019. **128**: p. 201-213.

[27] Choudhary, A. and A. Arora, *Linguistic feature based learning model for fake news detection and classification.* Expert Systems with Applications, 2021. **169**: p. 114171.

[28] Jehad, R. and S.A. Yousif. *Classification of fake news using multi-layer perceptron*. in *AIP Conference Proceedings*. 2021. AIP Publishing LLC.

[29] Singhal, S., et al. *Spotfake: A multi-modal framework for fake news detection*. in *2019 IEEE fifth international conference on multimedia big data (BigMM)*. 2019. IEEE.

[30] Wang, Y., et al. *Eann: Event adversarial neural networks for multi-modal fake news detection*. in *Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining*. 2018.

[31] Khattar, D., et al. *Mvae: Multimodal variational autoencoder for fake news detection*. in *The world wide web conference*. 2019.

[32] Tacchini, E., et al., *Some like it hoax: Automated fake news detection in social networks.* arXiv preprint arXiv:1704.07506, 2017.

[33] Kapusta, J. and J. Obonya. *Improvement of misleading and fake news classification for flective languages by morphological group analysis*. in *Informatics*. 2020. Multidisciplinary Digital Publishing Institute.

[34] Andreu, J., P.J.J.o.A.I. Angelov, and H. Computing, *An evolving machine learning method for human activity recognition systems.* J Ambient Intell Human Comput, 2013. **4**(2): p. 195-206.

[35] Liu, Q., et al. *Text features extraction based on TF-IDF associating semantic*. in *2018 IEEE 4th*

International Conference on Computer and Communications (ICCC). 2018. IEEE.

[36] Masadeh, S., J. Zraqou, and M. Alazab. *A NOVEL AUTHENTICATION AND AUTHORIZATION MODEL BASED ON MULTIPLE ENCRYPTION TECHNIQUES FOR ADOPTING SECURE E-LEARNING SYSTEM 1*. 2018.

[37] Park, S.-T., et al., *A study on smart factory-based ambient intelligence context-aware intrusion detection system using machine learning.* J Ambient Intell Human Comput, 2020. **11**(4): p. 1405-1412.

[38] Alweshah, M., et al., *Journal: Journal of Ambient Intelligence and Humanized Computing, 2019, № 8, p. 3405-3416.* 2019(8): p. 3405-3416.

[39] Stuart, A., *Kendall's advanced theory of statistics.* Distribution theory, 1994. **1**.

[40] Bremner, D., et al., *Output-sensitive algorithms for computing nearest-neighbour decision boundaries.* Discrete & Computational Geometry, 2005. **33**(4): p. 593-604.

[41] VenkateswarLal, P., et al., *Ensemble of texture and shape descriptors using support vector machine classification for face recognition.* J Ambient Intell Human Comput, 2019: p. 1-8.

[42] Zisserman, A., *Lecture 2: The svm classifier.* C19 Machine Learning (Hilary Term 2015). Available online at: http://www. robots. ox. ac. uk/~ az/lectures/ml/lect2. pdf (accessed March 23, 2019), 2015.

[43] Dorfman, R., *A formula for the Gini coefficient.* The review of economics and statistics, 1979: p. 146-149.

[44] Lin, Y. and Y. Jeon, *Random forests and adaptive nearest neighbors.* Journal of the American Statistical Association, 2006. **101**(474): p. 578-590.

[45] Pele, D.T., E. Lazar, and A. Dufour, *Information entropy and measures of market risk.* Entropy, 2017. **19**(5): p. 226.

[46] Hosmer, D.W., S. Lemeshow, and R.X. Sturdivant, *Applied logistic regression*. 2000: Wiley New York.

[47] Sahu, P. and R. Miri, *A Hybrid Technique for creating classification model using Random Committee and Voted Perceptron Classifier.* International Journal for Research in Applied Science & Engineering Technology, 2017. **5**(6): p. 2321-9653.

[48] Cohen, W.W., *Fast effective rule induction*, in *Machine learning proceedings 1995*. 1995, Elsevier. p. 115-123.

[49] Pan, X., et al., *Identifying patients with atrioventricular septal defect in down syndrome populations by using self-normalizing neural networks and feature selection.* Genes, 2018. **9**(4): p. 208.

[50] Alazab, A., et al. *Using feature selection for intrusion detection system*. in *2012 international symposium on communications and information technologies (ISCIT)*. 2012. IEEE.

[51] Olson, D.L. and D. Delen, *Advanced data mining techniques*. 2008: Springer Science & Business Media.

[52] Alazab, A., et al. *Web application protection against SQL injection attack*. in *Proceedings of the 7th International Conference on Information Technology and Applications*. 2011.

[53] Alazab, M., et al., *COVID-19 prediction and detection using deep learning.* International Journal of Computer Information Systems and Industrial Management Applications, 2020. **12**: p. 168-181.

[54] Batten, L.M., V. Moonsamy, and M. Alazab, *Smartphone applications, malware and data theft*, in *Computational intelligence, cyber security and computational models*. 2016, Springer, Singapore. p. 15-24.

[55] Falah, A., et al., *Improving malicious PDF classifier with feature engineering: A data-driven approach.* Future Generation Computer Systems, 2021. **115**: p. 314-326.

[56] Setiawan, R., et al., *Certain Investigation of Fake News Detection from Facebook and Twitter Using Artificial Intelligence Approach.* Wireless Personal Communications, 2021: p. 1-26.