International Journal of Computer Information Systems and Industrial Management Applications.
ISSN 2150-7988 Volume 14 (2022) pp.326-337
© MIR Labs, www.mirlabs.net/ijcisim/index.html

Received: 18 December, 2021; Accepted: 8 April, 2022; Publish 26 May, 2022

Fake News Detection and Prevention Using Artificial Intelligence Techniques: A Review of a Decade of Research

Albara Awajan[1], **Moutaz Alazab**[1], **Ruba Abu Khurma**[1], **Reem Alsaadeh**[1], **Mohammad Wedyan**[1] **and Ajith Abraham** [2]

[1]Faculty of Artificial Intelligence, Al-Balqa Applied University, Amman, Jordan
a.awajan@bau.edu.jo, m.alazab@bau.edu.jo, rubaabukhurma82@gmail.com reemsdh@gmail.com mwedyan@bau.edu.jo

[2]Machine Intelligence Research (MIR) Labs, Auburn, WA, USA
ajith.abraham@ieee.org

Abstract:
The circulation of fake news among people is not something new, as it has been present ages ago. In a connected world, due to the rapid development in the means of communication, fake news has become a very dangerous factor in daily life due to its massive impact. Furthermore, the size and speed of data shared through mediums makes it is difficult to differentiate fake and legitimate information. Social media allows sharing of data with low cost and easy access. This causes a harmful impact on individuals and society. Fake news classification and related topics has become an attractive topic for researchers in many disciplines such as journalism, natural language processing and national security . This paper reviews the various methods and techniques used in solving fake news problem and investigates weaknesses in the methods and techniques used in literature review.

The challenge is to find the most useful technology for detecting and mapping fake news. We concluded that many techniques - systems were designed and implemented to automate the process of detecting fake and misleading news, and also identified that deep learning techniques have a great ability to categorize and identify hidden correlation between multiple features in several fake news benchmark datasets in a way that overcomes human capabilities.

Keywords: Fake news detection, Machine learning, Deep learning, news verification, fact-checking, misinformation, information credibility

I. Introduction

The Fourth Industrial Revolution, 4IR is the result of increased interconnectedness and intelligent automation that has extended to cause rapid changes in all aspects of technological, industrial, and societal processes [Xu et al., 2018, Hasan et al., 2022,Batten et al., 2016]. Nowadays, social networking sites, including social platforms have become the main source of news and data sharing among one million users. There are 3.50 billion social media users worldwide. Social media [Win, k] goes beyond traditional media like TV and radio which has transformed traditional media into social media platforms [Shu et al., 2017]. Using social media, anything can reach millions of people around the world in a matter of hours. Traditional media, such as TV ads, newspapers, cold calling, and banner ads follow laws and regulations. All articles and news are checked for accuracy and integrity. On the contrary, social media is open to everyone and has no limits. This makes social media an opportunity for many people and agencies to unintentionally or even maliciously spread [Alazab et al., 2011b, Alazab et al., 2020a] and broadcast their rumors and fake news to the public [Conroy et al., 2015].

Dissemination of incorrect information among the people is not a new concept as it has been since ancient times and certainly before the creation of the Internet [Dawson, 2015, Alazab et al., 2011a]. Fake news is used to spread misleading and false information for self-interest. The spread of news has come a long way since it started as the news of the old day was spread through word of mouth from one person to another. At that time the spread of news was in the same area. After the revolution of communication and paper publications, the distribution of news is through news agencies that use radio, television and newspapers to broadcast news across the country and region. These agencies are used to monitor content and filter out any misleading, offensive or fake news [Alazab et al., 2021, Alazab et al., 2020a, Alazab, 2020].

After the uprising of the Internet and 4IR, the distribution and spread of news is easier than ever. Social media connected globally people and encouraged them to start spreading the news because it was so easy and free and could spread to billions of people all over the world. The fake news can have a terrible impact on people's personal lives, societies, politics. . . etc.

Online social platforms are beneficial to users because they can easily access news at little cost. But the problem, is that they are exploited by cyber criminals to spread fake news and promote it via these platforms. This news can be harmful to a particular person, group or political party [Alazab et al., 2012, Alazab, 2014, Alazab and Batten, 2015]. The danger is that this news is read, circulated and believed without validation. Detecting fake news is a big challenge because it is not an easy task. People's opinions and decisions are affected by fake news such as in the United States Elections 2016 [Tandoc Jr, 2019].

Exposing fake news recently has become listed as a national priority and a major concern of many governments and researchers [Al-Ahmad et al., 2021]. The wide spread of rumors, fake news and fake companies can have many dangerous effects on individuals and societies, such as racist ideas, fear, bullying, violence against innocent people and democratic influences. In addition to the many false stories associated with patients, their treatment and with many diseases such as cancer or diabetes, the adoption and belief of these false stories may lead to medical decisions and actions that harm the patient. Therefore, governments should take serious procedures and actions against anyone who misuses social media or at least alert people about fake or misleading news.

To manually check the accuracy and integrity of every article or every news published on social media, it will cost huge efforts for individuals and will cost a lot of money. Hence, there should be more focus on automated applications and techniques to detect and classify social media news as fake or real. With the number of fake news popping up every day, there is a need to have a model that can handle their classification with accuracy and perform well, at cheap cost and simple resources.

In the literature there have been many interests in developing new methods for detecting news and distinguishing fake from real. There is a clear trend towards applying machine learning algorithms that use supervised learning to identify and distinguish between fake and authentic news. In addition, a high percentage of fake news classification research has concentrated on the English language, with little attention being paid to other languages, such as Arabic. Arabic languages is considered a difficult language due to its structure and the presence of many local dialects rather than a single official language.

A summary of the main contributions provided in this survey is listed below.

- A classification for automatic fake news detection and mapping techniques is based on combining machine learning methods and news written language.
- Identifying the most useful techniques for detecting and classifying fake news taking into account the performance and accuracy measures.
- Emphasis and focus on how machine learning solutions are formulated to find accurate models for fake news classification and detection.
- The discussion of the advantages and disadvantages of each technology.
- The available datasets are classified and summarized.
- More specifically, the survey provides an overview of the recent studies that focused on fake news classification and detection and a comparison based on the used datasets, model building, and performance.

The sections of this survey are: Section II presents the general model of text detection. The public fake news datasets is presented in Section III. Section IV shows the preprocessing methods, Section V shows a taxonomy of features extraction methods. A taxonomy of fake news detection methods is presented in Section VI, an intense discussion and analysis of the selected studies is shown in Section VII, while Section VIII provides an overall conclusion of the survey work.

II. Text Detection General Model

Figure 1 shows the general model of text detection. The first step is the dataset collection. The second step is the preprocessing stage. Features extraction is performed in the third step. After that, in step four, the model is trained using a specific learning algorithm. The last step is the classification of the content based on the generated model. The content is classified either as fake or true using a binary classification. These sequential steps represent the general model for text detection where we discuss each step in detail the upcoming sections.

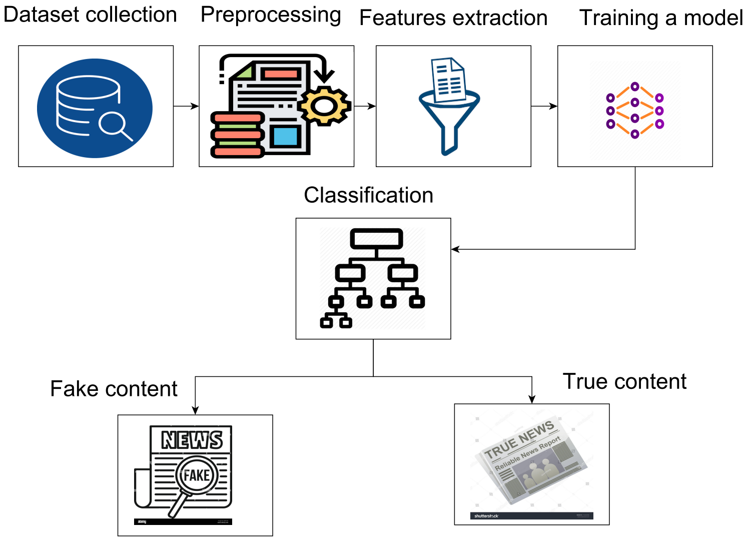

Figure. 1: Text detection general model

III. Public fake news datasets

The main difficulty that researchers confront when building a model for detecting and classifying of fake news in research is the quality and availability of the collected datasets. This work categorizes some of the related public fake news datasets into two categories based on the language to English datasets and Arabic datasets.

A. *English Fake news datasets*

The first public fake news dataset was released by Vlachos and Riedel [Vlachos and Riedel, 2014], they gathered data from two sources, the first one is POLITIFACT [Win, l] considered as Facebook and TikTok partner to help slow the spread of misinformation over the internet by flag all the posts that may have fake information or misleading. The second one is CHANNEL4.COM [Win, f] that has a fact-checking feature to seek the truth behind claims made by those in public office. This dataset contains 221 short claims; each one consists of three parts: the date it was made, the speaker, and a rating from five point score. [Oshikawa et al., 2018].

Another known dataset is the Emergent dataset [Ferreira and Vlachos, 2016] that was collected by journalists and contains

300 rumored statements with and additional 2,595 related news articles that were gathered from many online resources, such as rumor websites like Snopes.com [Win, e] and Twitter accounts like @Hoaxalizer. The throwback of this dataset is that it can improve the fact checking only when some articles related to the claim are given.

The main throwback of the datasets above is that both of them contain a small number of data so that they are impractical to be used for machine learning. One of the recent fake news detection datasets is LIAR [Wang, 2017] which was collected by Wang, he gathered around 12.8K short statements that were labeled manually with six-grade truthfulness from POLITIFACT5. Some datasets were generated from Wikipedia such as Fever dataset [Thorne and Vlachos, 2018] which contains 185,445 statements.

Many datasets that are associated with fake news and reviews detection can be found such as FakeNewsNet [Shu et al., 2017] and BS DETECTOR [Win, h]. The FakeNewsNet dataset includes data gathered from two fact-checking platforms, such as: BuzzFeed and PolitiFact, each article contains a section for the headlines and a section for the body texts of fake news articles. BS DETECTOR dataset uses the BS Detector browser extension to collect any URL on HTML web page that links to a questionable reference and checks news veracity.

FA-KES dataset [Salem et al., 2019] is a dataset that includes fake news about the war in Syria. The dataset contains 804 news articles labeled as true or fake from many media outlets that represent the pro-government press, mobilization press, and diverse print media. In their work, they avoid the difficult task of manually labeling by labelling the news articles in the dataset as fake or not using the semi-supervised fact-checking approach.

Deceptive Opinion Spam Corpus (DOSC) dataset [Ott et al., 2011] is a large-scale spam dataset that holds misleading opinions. The DOSC utilizes the TripAdvisor site [Win, m] that uses a proprietary ranking system to assess hotel popularity to collect the truthful opinion. They chose the 20 highest ranking hotels irrespective of the TripAdvisor ranking. In addition to 400 false reviews that were generated by human-intelligence tasks (HITs) by Amazon Mechanical Turk (AMT) [Win, b]. For the same chosen hotels they gathered 20 truthful and 200 false opinions for each of the 200 chosen hotels (800 opinions total).

B. Arabic fake news datasets

COVID-19-FAKES dataset [Elhadad et al., 2020a] includes data related to COVID-19 that was collected for around 36 days starting from February 04, 2020. The authors gathered information and labeled them as reliable or fake by checking them against trusted and authorised website such as the WHO, UNICEF, and UN formal websites. The Authors worked on building a solid database for reliable information by using different fact-checking sites. Data was collected in real-time with the use of an API streaming options, and that have some keywords related to the COVID-19, such as (Coronavirus, Novel, COVID-19, Corona-virus, Coronavirus, Corona, virus, etc). The dataset consists of 3.263M English and Arabic Tweets that are classified as real or Misleading.

Arabic Corpora dataset [Al Zaatari et al., 2016] contains two Arabic corpora for credibility analysis, one of them consisting of 2708 Tweets and one of 175 Blog posts. In their work, they built machine learning models using an exhaustive list of features to verify the usefulness of their corpora. Their annotated corpora are the first of their kind and they will be a valuable resource for future studies on the credibility of Arabic content.

ArCOV19-Rumors dataset [Haouari et al., 2020] collected from 27th January in 2020 and continued for three months and includes 18 true confirmed claims and 113 confirmed false claims and over 9K tweets that are directly associated to the claims. The dataset ArCOV19-Rumors dataset consists of two classes of fake news detection on Twitter: claim-level verification which is verifies free-text claims and tweet-level verification which verifies claims that are expressed in tweets. They collected their claims from Fatabyyano [Win, g] and Misbar [Win, j] which are two popular Arabic fact-checking websites that are specialized in the field of verifying news. The ArCOV19-Rumors dataset mentioned earlier expands from ArCOV-19 [Haouari et al., 2020] which is considered the first Arabic Twitter dataset that is specialised in COVID-19 and includes one million Arabic tweets. The authors manually-annotated the tweets in order to to give support for both tweet-level verification and claim-level tasks.

IV. Preprocessing Methods

This step is essential for removing and reducing the noises and the unwanted parts from data before extracting any feature to improve the classification performance. Many processes attracted several researchers, for example tokenization, normalization, removing of stop words, and light stemming. Fig 2 shows the prepossessing steps.

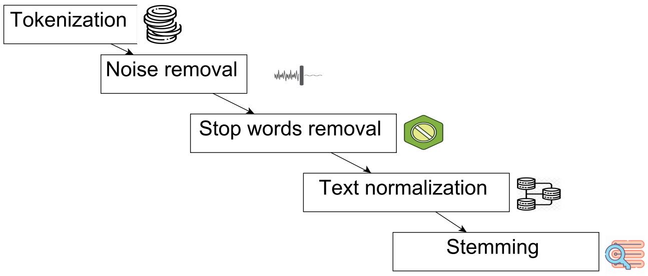

Figure. 2: Preprocessing steps

- **Tokenization** The tokenization method is used for splitting the text into smaller components. Each component is called a token, and each token is expressed by a single word based on a white space character. Many natural language processing tasks need access to each word in the text to provide many operations, such as: determining how many words appear in the sentence. How many sentences appear in the paragraph and counting the occurrence of a specific word in the defined text.
- **Noise Removal** Based on the project goal, you may need to remove unwanted information in the data by checking all the text tokens. For example: removing the special characters, removing the numeric digits,

removing HTML formatting, and removing the vertical whitespace. In addition, for Arabic text detection projects; you can remove all the non-Arabic characters to reduce the data size.

- **Stop Words Removal**
Stop words are the widely used words in the language and are meaningless that will make no difference in text detection and classification. Removing these words can improve the result by reducing both the response time and the index's space. The English language has a lot of stops words, such as a, an, about, are, is, as, at, for, from, at, be, by, that, the, too, was, were, and so on. Also, the Arabic language uses around 700 stop words in sentences to help connect the sentence structure.
- **Text Normalization** Some text needs processing through text normalization which aims to unify different words into a common form by changing the input text into a consistent output. This step is usually used after the tokenization pre-processing method, by checking all characters in each token to detect if they are normalized or not. Several English text detection researchers used the normalization process by converting the uppercase to lowercase or vice versa, on the hand Arabic text detection researchers used it by replacing some letters into one letter to produce one form of the characters, see table1.

Table 1: Example of normalizing Arabic text

Letters to change	Changed with
ى ، ىٔ ، ئ	ي
آ ، أ ، آ ، أ ، إ	ا
ة ، ه	ه
ؤ ، ؤ ، ؤ	و

- **Stemming** The act of returning the word to its standard origin form, the main objective of the stemming is minimize the overall word classes or types in the data. For non-Arabic languages, the stemming is concerned with removing word affixes (prefixes and suffixes) to express a grammatical syntax, for example, cast the word "going" to "go" or by returning the word to its original root, like: "work" ,"worker", and "working" will be cut/minimized to the original word "work". In the Arabic language, stemming is a more difficult process because several words after the stemming process will have the same root although they have different meanings. so to solve this problem; the Light Stemming process can be used instead of the Stemming process. Light stemming can be used to crop the prefixes and suffixes from the Arabic word.

V. Taxonomy of Features Extraction

The features extraction techniques are an important phase because selecting the proper features will directly enhance the detection results. One of the challenging factors of the text classification is the high dimensionality of data. Therefore, using the feature reduction is very important to reduce the number of features. There are many features extraction techniques. However, this work, summarizes the most methods used in the literature review, such as N-gram feature extraction, Term Frequency–Inverted Document Frequency (TF-IDF) and Negation Handling. Figure 3 shows the classification of features extraction methods.

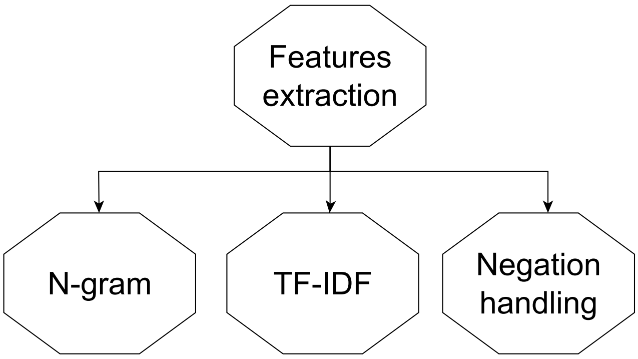

Figure. 3: Features extraction methods

- **N-gram Feature Extraction** The n-gram model is a common feature analysis and identification method. This method is used intensively in Natural Language Processing (NLP) Language modeling domain and is a popular method for text classification. The N-gram consists of a stream of N adjacent items where an item could be a stream of characters or words and this is the most common; or it can be syllables, or bytes. The word-based N-grams are divided to three main classes based on the number of contiguous words: Uni-gram, Bi-gram, and Tri-gram.
- **Term Frequency–Inverted Document Frequency (TF-IDF)** The TF-IDF is a weighted metric mostly used in NLP and information retrieval. The working principle of the term frequency technique is utilizing word counting redundancy in documents. an increase in word count indicates the importance of that word in a document and the importance is reflected to the dataset also.
- **Negation Handling** The negation handling approach is usually used after the N-gram model. It depends on the language negation words. In this step, all words prior to the extracted N-gram features are examined. This helps in detecting of the word is a negation word or not to correct the polarity classification. (no, not, no one, nobody, nowhere, etc.) are examples of the English negators. The Arabic language has a list of 50 negators words, such as (ليس ، لا ، لم ، مش ، لن ، مو).

VI. Fake news detection approaches.

This section presents a systematic literature review for fake news detection. Furthermore, it shows the selected way that is used to filter papers and analyze the research contributions in the papers, as well as, the results. We selected to lter the research based on two phases, the fake news language which includes the Arabic or English Language, and the detection method for fake news which includes Machine Learning methods (ML) or Deep Learning methods (DL).

A. Detection Using Machine Learning Methods

ML is an application of Artificial Intelligence (AI) where ML applies various algorithms on datasets in order to teach computers to take actions without explicitly being [Win, i]. The four main learning methods: Supervised, Unsupervised, Semi-Supervised and Reinforcement Learning [Chapelle et al., 2009]. In machine learning, detection goes under supervised learning where the set contains the correct output labels. Many algorithms have been dedicated to solve detection problems and are known as classiers, such as Naïve Base [McCallum et al., 1998], Decision Tree [Geurts et al., 2005], Multi-Layer Perceptron (MLP) [Lippmann, 1989], and Support Vector Machine (SVM) [Guyon et al., 2002]. There have been many research works about fake detection in the literature. They give attention to machine learning techniques that uses supervised learning to recognise misleading and fake news from genuine news..

1) Arabic Fake News Detection Using Machine Learning

As the majority of researchers concentrated on the English language for their work on fake detection [Faustini and Covoes, 2020]. Other languages did not gave the attention they deserve. Although there is a large growth in Arabic content in social media and on the Internet, but there are not many studies that shed the light on Arabic fake news classification. Authors in [Saeed et al., 2022] used four different methods to detect Arabic spam reviews: Machine Learning Classifiers, Rule-based Classifier,such as Naive Bayes, Logistic Regression, Support Vector Machine, and Decision Tree. Furthermore, Stacking Ensemble and Majority Voting Ensemble Classifiers.

Authors in [Saeed et al., 2022] used three main modules as a proposed approach to detect spam reviews written in Arabic. rst, they used the Pre-processing technique that went through a five-stage process: normalization, discarding of stop words , tokenization, discarding of non-Arabic text, and light stemming. This module helped in getting rid of all noisy parts of data before going into the next module that will extract appropriate features. Then they used a three process Extraction Module to extract the features directly from the data: first, N-gram Feature Extraction. second, Negation Handling. And third, Content-based Feature Extraction.

In their study, the four methods were applied on to two datasets. The first dataset consists of hotel ratings and reviews in the city of Chicago (DOSC) [Win, b]. The dataset contains a total of 1,600 reviews that were translated to English. The other applied dataset was the ”Hotel Arabic Reviews Dataset” (HARD) [Elnagar et al., 2018] where the data was collected from around 1800 hotels listed in Booking.com. The data set holds 94,052 reviews written in Arabic. Unfortunately, the main throwback is that the reviews in the dataset were not labeled as fake or not. The best accuracy obtained by this research is 99.98% by using the Stacking Ensemble method. The main pros of this research are that they used the Stacking Ensemble that executed the classifiers sequentially and not in parallel.

Authors in [Wahsheh et al., 2013] developed an Arabic spam detection system URL. The system classifies the reviews on Yahoo Maktoob social network by testing several features in to either spam or not spam. Then the spam review is checked against distinct parameters and categorized into high-probability or low-probability spam. Then the review is checked with the language polarity dictionaries and depending on the check it is labeled as spam, not spam, or neutral. If the URL to be tested is not labeled as spam or the link has the symbol ‘@’ with consecutive letters or numbers then the review will be added to the low-level spam class. On the other hand, if the a URL is not part of the review, then it will be add to the non-spam class. They used the Support Vector Machine (SVM) classier and a 97.5% accuracy was obtained.

Authors in [Abu Hammad, 2013] merged different techniques from text mining and data mining into one mining classification approach (SDAOR) that was implemented by using RapidMiner; RapidMiner a machine learning tool that introduces a new approach for detecting the Arabic opinion reviews, their work is based on the Latin-based spam detection techniques but they put into account the special characteristics of the Arabic language, they collected the data from three sites, including TripAdvisor [Win, m], Booking [Win, d], and Agoda [Win, a] using a crawler to obtain TBA dataset where they combined different methods such as meta-data, review content features, and reviewer features. The authors evaluated the system by using Support Vector Machine, K-Nearest Neighbor and Naive Bayes where the later obtained 99.2% accuracy.

Authors in [Sabbeh and BAATWAH, 2018] did their research on Twitter where they designed a model for classifying Arabic fake news, they used a hybrid set of features to evaluate news credibility, their work consists of four main modules: features extraction and Content parsing module, content Verification module, users’ comments polarity evaluation and credibility classification module. They collected 800 Arabic news that is manually labeled as a dataset from Twitter where they utilized content-based, user-based, and sentiment analysis features in their work. They claimed that significant increase in performance can be obtained by using sentiment analysis features. The authors trained and tested their models on different machine learning techniques, such as Decision tree, support vector machine (SVM) and Naive Bayesian(NB), where the highest obtained accuracy was 89.9% by the Decision Tree classifier. In our point of view, the drawback of this research is that they used a small dataset; we think if they augment the data, they will improve the accuracy.

Authors in [Jardaneh et al., 2019] a model that detects the Arabic fake news on Twitter. The model used the content-based, user-based, and sentiment analysis features where 45 features where extracted for each tweet. As [Sabbeh and BAATWAH, 2018] concluded that using the sentiment analysis had a positive impact on enhancing the prediction model, they choose four machine learning classifiers to train and test their models namely Random Forest, Decision Tree, AdaBoost, and Logistic Regression, they applied thier work on a dataset that contains 1862 tweets that focused on the Syrian war. The Random Forest classifier scored an accuracy of 76% . The main cons in their work are using a high number of features to detect the Arabic fake news that decreased the model performance.

Authors in [Alorini and Rawat, 2019] proposed a system that classifies the Gulf Dialect Arabic tweets as spam or not spam, they differentiated between legitimate and illegitimate users by studying both user and content attributes, they used an API to collect 2000 tweets that includes: user ID, hash-tags, URLs, and ID's of retweets, then they normalized the tweets by removing non-Arabic words but kept the hash-tags and all additional words were discarded such as emphasizing words which resulted in full Arabic text. The authors extracted several features in their work such as shortened URLs, hashtags, and presence of profanity words. The system was evaluated by using two different classifiers: support vector machine and Naive Bayes, where a the later classifier scored max accuracy with a percentage of 86% .

In other studies, researchers used semi-supervised learning and unsupervised learning models in their systems to reveal Arabic fake news using expectation-maximization (E–M), such as [Alzanin and Azmi, 2019], that collected a total of 271,000 tweets using search API and from the anti-rumors authority, they gained the rumors topics [Win, c]. They extracted a set of features from the dataset, then they analyzed these features to measure their significance. They compared between supervised Gaussian Naïve Bayes (NB) and semi-supervised learning; they concluded that their semi-supervised learning system outperforms Naïve Bayes by obtaining a 78.6% accuracy.

Table 2 shows Different machine learning methods for Arabic fake news detection.

2) English Fake News Detection Using Machine Learning

Authors [Mani et al., 2018] proposed an algorithm that combines many learning algorithms for better predictive to detect fake news by using an a technique that merges the three classifiers: 1-Random Forest, 2-Naive Bayes, and 3-Support Vector Machine. The authors used the N-gram features extraction to extract the features from the DOSC dataset which contains 1,600 hotel rating reviews for around 20 hotels in the Chicago region and obtained a 87.68% accuracy. This research proved that when applying simple features to their work such as the ensemble method and the N-gram method can enhance the performance and accuracy of fake news classification.

Authors in [Saumya and Singh, 2018] proposed a novel model for spam detection that used three features, such as sentiments of review and the related comments, content-based factor, and rating deviation. The data set they gathered was from Amazon where just beneath 40k reviews for online bought products was collected using the scrapper software. They used the Random Forest classifier and they achieved an F1 score of 91%. In another work, authors [Anil Kumar et al., 2018] applied semantic and machine learning algorithms: Decision Trees, Nearest Neighbor, Artificial Neural Network (Multilayer Perceptron), Naive Bayes, SMO, and Logistic Regression on five datasets by using the Weka tool. They applied their model on the collected datasets that included spam and non-spam reviews. The authors conclude that the machine learning approach outperforms the semantic approach where their model performance was 82.2% when applying a Neural Network classifier.

Authors in [Hassan and Islam, 2019] introduced semi-supervised and supervised text mining models for fake reviews detection, they used the content-based features including length of review, word frequency count, and sentiment polarity. The authors used the Gold Standard dataset developed by Ott15, where the dataset included 1,600 rating reviews for hotels in the Chicago region, USA. They applied the Support Vector machines(SVM) and Naive Bayes(NB) classifier with EM algorithm as a classifier. The evaluation of the model showed that the Naive Bayes classifier obtained the highest performance with a score of 86.32%. Authors in [Narayan et al., 2018] applied a supervised learning technique and built models by using a different set of features, such as The Linguistic Inquiry and Word Count (LIWC), parts of speech (POS) Tags, N-Gram Feature, and Sentiment Score feature to detect English spam review. As [Hassan and Islam, 2019] summarized above, they used the Gold Standard dataset developed by Ott [Ott et al., 2012], they employed Six classification algorithms, such: As naive Bayes, SVM, k-NN, random forest, and logistic regression. The Authors combined the overall features of LIWC and unigram to result in an accuracy of 86.25% ;

Ahmed [Ahmed et al., 2018] proposed a new n-gram model that used to automatically detect fake spam and fake news contents, they used two different features extraction techniques, namely, term frequency (TF) and term frequency-inverted document frequency (TF-IDF), and they applied 6 machine learning classification techniques, namely: stochastic gradient descent (SGD), support vector machines (SVM), linear support vector machines, K-nearest neighbor, logistic regression (LR) and decision tree (DT), on two datasets, first one collected by Ott [Ott et al., 2012] that contains 1600 true and fake English reviews. and the other dataset collected by the authors contains 12600 fake news article and 12600 truthful articles that collected from real-world resources, for truth opinion the collected the news article from Reuters.com, and for fake news, they used the fake news datasets on the Kaggle), they used a 5-fold cross-validation, and they split the dataset in each validation round to 80% for training and 20% for testing. Their results show that the linear-based classifiers (Linear SVM, SGD, LR) obtained better results than the non-linear ones, and when increasing the n-gram to tri-gram and four grams the accuracy will decrease. The best accuracy was 92% and obtained by linear SVM.

Elhadad [Elhadad et al., 2020a] proposed a model to distinguishes misleading information with relation to COVID-19. The model gathers data from the World Health Organization, UNICEF, and the United Nations. They constructed a voting ensemble machine learning classifier by using Ten machine learning algorithms (DT, MNB, BNB, LR, KNN, Perceptron, NN, LSVM, ERF, XGBoost), with seven feature extraction techniques (Term Frequency, Unigram, Bigram, Trigram, N-

Table 2: Comparison Between the Arabic Fake News Detection Using Machine Learning Literature Review

Source	Dataset	Method	Classifier	Feature Extraction	Accuracy
[Saeed et al., 2022]	(DOSC) (HARD)	using the Stacking Ensemble that execute the rule-based classifier and machine learning classifiers sequentially	1-Rule-based Classifier. 2- Machine Learning Classifiers, such as: DT, NB, LR, SVM, K-Means, KNN, Bagging, Boosting, RF and NN. 3- Majority Voting Ensemble Classifier. 4- Stacking Ensemble Classifier.	1-N-gram Feature Extraction 2- Negation Handling 3- Content-based Feature Extraction	99.98%
[Wahsheh et al., 2013]	collection of Modern Standard Arabic (MSA), and colloquial Arabic opinions (reviews)	developed an Arabic spam URL detection system that classified the reviews to spam and non-spam based on the number of features	SVM	Term frequency (TF)	97.5%
[Abu Hammad, 2013]	TBA dataset that collected from online Arabic economic websites: tripadvisor.com.eg, booking.com, and agoda.ae	proposed a new approach for detect the Arabic opinion reviews by merged methods from data mining and text mining into one mining classification approach (SDAOR)	SVM KNN Naive Bayes	1-Review Content 2- Meta-data about each Reviewer 3- Product Information	99.2%
[Sabbeh and BAATWAH, 2018]	Collected 800 Arabic news from Twitter	proposed a model for detecting fake Arabic news by using a hybrid set of features to evaluate news credibility	Decision Tree SVM Naive Bayes	Hybrid features including client-based features (web application client and mobile client program type) and location-based feature	89.9%
[Jardaneh et al., 2019]	Arabic Corpora dataset	Proposed a model to that detect the Arabic fake news by using sentiment analysis	1-Random Forest 2-Decision Tree 3-AdaBoost 4-Logistic Regression	1-user-based 2-content-based 3-sentiment analysis	76%
[Alorini and Rawat, 2018]	Collected around 2000 Gulf Dialect Arabic tweets	Introduced a model to differentiated between legitimate and illegitimate users by studding both user and content attributes	SVM Naive Bayes	1-number of hash-tags 2-number of shortened URLs 3-existence of profanity words	86%
[Alzanin and Azmi, 2019]	collected A total of 271,000 tweets using search API	used two different learning models to detect Arabic fake news; semi-supervised learning and unsupervised learning using expectation–maximization (E–M)	Naïve Bayes (NB) semi-supervised learning	1-user-based 2-content-based	78.6%.

gram, Characters Level, Word Embeddings), they used the COVID-FAKES [Elhadad et al., 2020a] dataset that contains 3,047,255 COVID-19 related tweets, they applied some of the pre-processing techniques on the dataset such as Text Parsing, Data Cleaning, Part of Speech (PoS) Tagging, Stop Words Removal, Stemming, etc. They performed a 5-fold cross-validation to check the validity of the collected data and they evaluated their model by using twelve performance metrics (Accuracy, Error Rate, Precision, Sensitivity, F1-Score, Specificity, Area Under the Curve, Geometric-Mean, Miss Rate, False Discovery Rate, False Omission Rate, and FallOut Rate), the best accuracy that obtained by this research is 99.80% when using the Term Frequency feature extraction with NN classifier.

As illustrated above, Arabic language is considered a complex language due to its vocabularies and numerous linguistic bases and grammar. Therefore, there are many features that algorithms can use to improve the performance of the Arabic fake news classification such as Stemming algorithms that is used to reduce words to their three-letters roots, Chi Square (CHI) and Information Gain (IG), contrary to the English language that needs simple features extraction techniques. On the other hand, most of the literature review adopted the n-gram technique as a feature extraction model rather than Unigram when working with Arabic language for better feature extraction.

Table 3 shows a comparison between English Fake News Detection Using ML.

B. *Detection Using Deep Learning*

A new branch of machine learning has appeared in late 2012 called Deep Learning (DL) [Alazab et al., 2020b], it is a class of machine learning methods that extracts features by using multiple layers of nonlinear processing units. DL represent data as levels of abstraction then provides these abstraction to the computational models that are composed of several processing layers to simulate the process of learning. Deep learning has many applications such as: sounds detection, text detection, image recognition, processing of Natural language and bioinformatics, which provides high value services that are regarded essential to our daily life.

One of the most popular and powerful deep learning models is convolutional neural network (CNN). CNN is a nonlinear and a complex kind of ANN compared to other methods because it contains many hidden layers. CNN is widely used in solving various detection problems. It also has many advantages such as: hardness to distortion in the image, fewer memory requirements and easier training [Hijazi et al., 2015]. In addition, CNN is a flexible computational tool that can handle large number of datasets. It can implicitly detect complex nonlinear relationships between dependent and independent variables with a high degree of accuracy.

1) *English Fake News Detection Using Deep Learning*

Authors in [Barushka and Hajek, 2019] proposed a robust model that focuses on a content based approach, they built a vector model by utilizing n-grams and the skip-gram word embedding method. They used a deep feed-forward neural network as a second step to identify the spam and not spam reviews. The authors applied their work on two hotel review datasets that were taken from Cornell University where one of the datasets included the positive reviews and the other contained the negative review, their model obtained 89.75% accuracy. Authors in [Jain et al., 2019] proposed the Convolutional Neural Network model (CNN-GRU) and the Multi-Instance Learning model (MIL). They applied their models on three different benchmark datasets: DOSC [Ott et al., 2011] that included 1600 reviews divided equally into genuine and fake reviews with positive and negative sentiments. Four-City [Li et al., 2013] dataset that cleary from its name is a review for eight hotels in four different cities, where each hotel has 40 real and 40 fake reviews. YelpZip [Rayana and Akoglu, 2015] dataset that includes 608598 reviews with around 80k false reviews and around 53K true reviews. They evaluated their proposed models by using two additional datasets: Large Movie Review dataset (LMRD) [Maas et al., 2011] that consist of 50,000 reviews from IMDB and Drug Review Dataset (DRD) [Gräßer et al., 2018] that contains

Table 3: Comparison between the English Fake News Detection Using ML

Source	Dataset	Method	Classifier	Feature Extraction	Accuracy
[Mani et al., 2018]	DOSC dataset	proposed an algorithm that combine many learning algorithms for better predictive for detect the fake news by using ensemble technique	ensemble technique combining: NB, RF, SVM	the n-gram (unigram and bigram)	87.68%
[Saumya and Singh, 2018]	Collected around 39,382 online product reviews by authors	proposed a novel and robust, spam review detection system, they address the aforementioned limitations by investigated all these features for only suspicious review list	Random Forest Gradient Boosting SVM	1-sentiments of review 2-content-based factor 3-rating deviation	91%
[Anil Kumar et al., 2018]	Authors collected five different datasets	using semantic and machine learning algorithms on five different datasets to classifying the product reviews into spam or non-spam	Decision Trees K-NN ANN Naive Bayes, SMO LR	sentiment analysis	82.2%
[Hassan and Islam, 2019]	DOSC dataset	introduced semi-supervised and supervised text mining models for fake news detection	SVM Naive Bayes (NB) classifier with EM algorithm	content-based including: 1-word frequency count 2-sentiment polarity 3-length of review	86.32%
[Narayan et al., 2018]	DOSC dataset	applied a supervised learning technique and built models by using different set of features	Naive Bayes SVM k-NN random forest logistic regression	1-The Linguistic Inquiry and Word Count (LIWC) 2- parts of speech (POS) Tags 3- N-Gram Feature 4-Sentiment Score feature	86.25%
[Ahmed et al., 2018]	1-DOSC dataset 2-dataset collected by authors that contains 25,200 articles	proposed a new n-gram model that used to detect automatically the fake spam and the fake news contents	SGD SVM linear SVM K-NN logistic regression (LR) decision tree (DT)	1-term frequency (TF) 2-term frequency-inverted document frequency (TF-IDF),	92%

215063 reviews from drugs.com website.

Elhadad [Elhadad et al., 2020b] proposed a model that classifies data related to COVID-19 that uses several deep learning methods and depends on the World Health Organization, UNICEF, and the United Nations a data reference. In order to enhance the total performance of their system, they implemented a features engineering preprocessing step. They used the COVID-FAKES18 dataset that contains 3,047,255 COVID-19 related tweets, they applied some of pre-processing techniques on the dataset such as: Text Parsing, Data Cleaning, Part of Speech (PoS) Tagging, Stop Words Removal, Stemming, etc. They constructed a voting ensemble Deep Learning classifier by using m 6 DL techniques (Sequential model, CNN, Recurrent Neural Network (RNN-LSTM, and RNN-GRU), Bidirectional Recurrent Neural Network (BiRNN-GRU), and Recurrent Convolutional Neural Network (RCNN)), they split the data to 80% for training and 20% for testing the model, and they used the sigmoid function as an activation function in the output layer. They evaluated their model by using 14 performance metrics (Accuracy, Error Rate, Loss, Precision, Recall, F1-Score, Area Under the Curve, Geometric-Mean, Specificity, Miss Rate, Fall-Out Rate, False Discovery Rate, False-Omission Rate, and the Total Training Time). To the best of our knowledge, there is still no published study to detect the fake Arabic news using deep learning techniques and CNN algorithm because it's a new topic and the Arabic language is more complex than English. This means that Arabic fake news detection is still in its early stage.

Table 4 shows a comparison between different architecture of deep learning models.

VII. Discussion and Analytical view of the selected studies

This section shows some analytical aspects of the studies papers and discusses the main outcomes from the conducted literature review. A summary of the authors and the category of the research is provided in Figures 4 and 5. Figure 4 shows the distribution of studies into three categories: Journal articles, websites and conference papers. The largest ratio which is 38.8% which represents the journal articles. The second ratio is conference papers which consists 34.7%. The smallest ratio is websites which consists 26.5%.

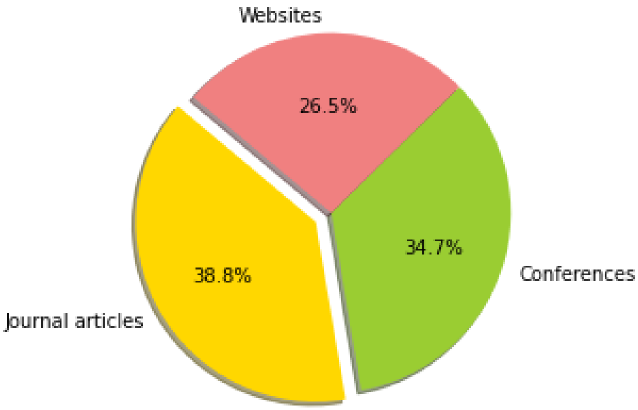

Figure. 4: Type of selected studies that described in this work

Figure 5 shows that the papers published in journals by publishers like Elsevier, then Springer, IEEE, ACM, MDPI and Hindawi.

Figures 6 and 7 conclude the extracted the relevant information from the research papers' database. Figure 6 shows the relation between the country and the number of published papers. It can be noticed that China has the highest number of published papers with 144 articles, then India with 93 articles, then USA with 82 articles. Canada published 76 articles, then Iran published 44 articles. Korea published 19 articles then Australia published the smallest number of articles which is about 12 articles.

Additionally, Figure 7 indicates the most frequently occurring keywords for these articles . It is obvious that the prominent keyword are 'Fake news detection', 'machine learning', followed by 'social media' and 'Deep learning'. On the other hand, it can be noticed that the least occurrence keywords are 'News verification', Veracity assessment', and 'Automation'. In summary, fake news classification and detection model is based on a set of systematic steps including dataset collection, applying preprocessing methods, adopting feature extraction techniques, training a model using a learning algorithm. Finally, is the classification of news into truthful news and fake news using a specific classification technique. In

Table 4: Comparison Between Different Architecture of Deep Learning Models

Type of Network	Detail of Network	PROS	CONS
Deep Neural Network (DNN)	Allows complex non-linear relationship because it consist of more than two layers	Usually achieves high accuracy	The learning process of the model is too much slow
Convolutional Neural Network (CNN)	consists of convolutional filters which transform 2D to 3D,	Very good performance, it is good for two-dimensional array	Needs a lot of labeled data for classification
Recurrent Neural Network (RNN)	The weights are sharing between all neurons and steps	Has many versions so it can provide many NLP tasks, such as: speech recognition and character recognition with high accuracy	Needs a big dataset
Deep Boltzmann machine (DBM)	It consists of unidirectional connections between all hidden layers. It is based Boltzmann family	The top down feedback in this model combine with ambiguous data so it is increasing the robust inference	The optimization of the big data is not possible
Deep Belief Network (DBN)	It is used with supervised and unsupervised learning. It has unidirectional connections at the top two layers. Each hidden layer in the sub-network is visible to the next layer	Each layer uses the greedy strategy	The initialization makes the training process more expensive
Deep Autoencoder (DA)	It is designed for extraction of dimensionality of features. It is used in supervised learning. The number of inputs in this network is equal to number of outputs	It doesn't need a labeled data	It needs pretraining step

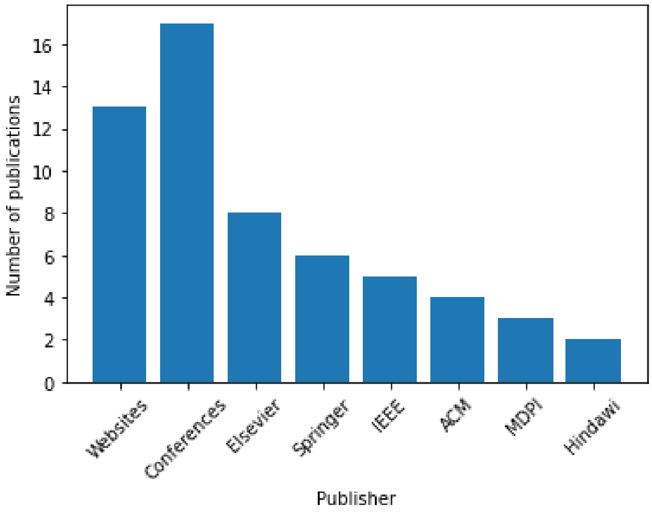

Figure. 5: Number of published papers vs publisher

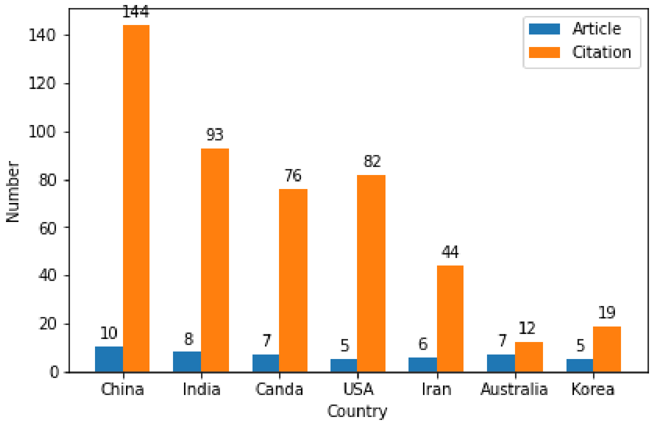

Figure. 6: Articles and citations per country vs country origin of revised papers

this study, the public datasets related to fake news classification were separated and grouped the based on the language to Arabic datasets and English datasets. The details of these datasets are discussed in Section III.

The preprocessing methods are the second aspect that should be taken into consideration when trying to expose the fake news and distinguish them from the true ones. Five primary steps have to be applied on the dataset including tokenization, noise removal, removal of stop words, text normalization and stemming. The details of preprocessing methods are discussed in Section IV.

Features extraction is an essential step for fake news detection as it plays a major role in dimensionality reduction of a dataset and keeping the most informative and relevant features that can enhance the overall performance of the classification process. The details of features extraction methods are presented in Section V.

The last step is the classification process for the text either it is fake or real. There are two major approaches applied for handling this job either by implementing machine learning approaches or deep learning approaches. The machine learning approaches are applied for both Arabic and English fake news detection. However, the deep learning approaches are tailored towards the English fake news detection. Until now, there is no deep learning approach developed to tackle the Arabic fake news detection. Section VI discusses the fake news detection approaches in detail. Table 5 summarizes the main differences between machine learning and deep learning approaches.

VIII. Conclusion and future works

The Internet is an invention that encompasses all aspects of life. The internet is used by billions of people around the globe. People use the Internet for various reasons where social media platforms are gaining very high popularity. Any user can join a platform and create a post or publish news without verifying user, post or content. Therefore, it is a fertile environment for the dissemination fake news. This fake news can target individuals, organizations or political parties.

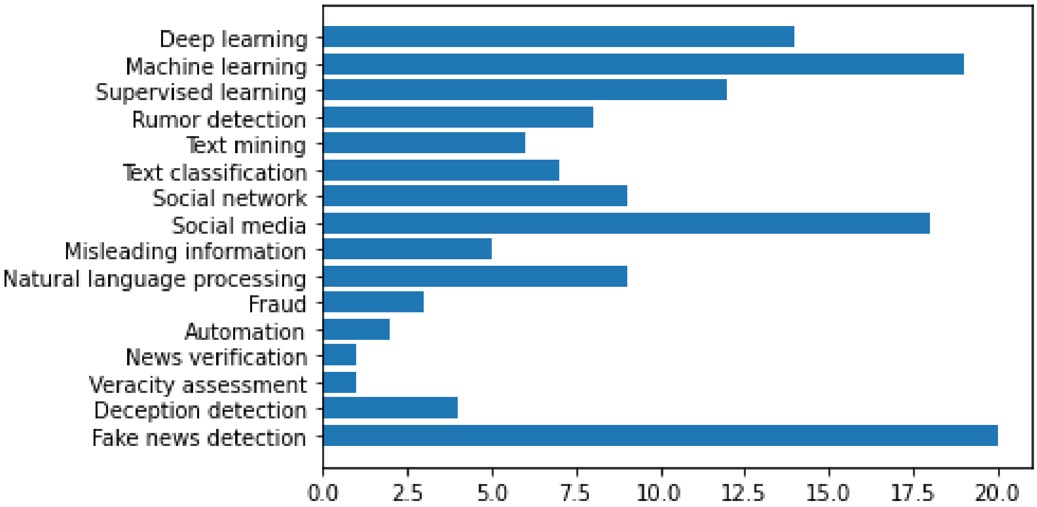

Figure. 7: Keywords vs Keyword Occurrences.

Table 5: Comparison between Machine Learning and Deep Learning

Factor	Machine Learning	Deep Leaning
Size of Data Set	Can train on lesser data	Requires large data
Training Time	Takes less time to train (needs a couple of minutes to hours)	Takes longer time to train (weeks and months)
Hardware Requirements	Trains on CPU	Trains on GPU and Requires large amounts of processing power
Hyperparameter Tuning	Limited tuning capabilities	Can be tuned in various different ways
Accuracy	Gives less accuracy	Provides high accuracy
Methodology	Rule-based, data driven process	Rule-based, data driven process that utilizes neural networks
Data Requirement	Uses Labeled data and labeled features	Uses unlabeled or unstructured data

One cannot detect all this fake news. Therefore, there is persistent need to design and implement automated fake news classification and detection methods.

In this study, first we summarized the definitions and importance of automatic fake news detection. Then we discussed and compared the most recent and most used benchmark datasets, and then we classified them based on the language. Then we discussed experimental results of different classification and detection methods. Based on our observation, the most high-performance model is the convolutional neural network (CNN) in deep learning. Furthermore, based on the literature review, using Ensemble models obtained better accuracy than using one machine learning classifier.

Due to the rapid development of fake and misleading news research, we included an online repository ”http://fake-news.site” along with the survey to provide continuous timely summaries and new applications for fake news, including educational programs, publications, new methods, datasets, and any other relevant resources.

For future, there are several research directions that can be performed including developing a dataset for fake news detection that can collect data from multimedia resources such as audio, video and images. The dataset can also be designed to be multi-lingual where it can combine different languages and compare different news in different regions. In addition it should be adaptive, cross-domain, and large-scale.

Acknowledgements and Fundings

The research reported in this publication was funded by the Deanship of Scientific Research and Innovation at Al-Balqa Applied University, Al-Salt, Jordan. (Grant Number: DSR-2020 #228).

Conflict of interest

The authors have no conflicts of interest to declare that are relevant to the content of this article. There are no conflicts of interest relevant this article to be declared by the authors

References

Agoda official site. `https://www.agoda.com/ar-ae/`. Accessed: 2022-04-17.

Amazon mechanical turk. `http://mturk.com`. Accessed: 2022-04-17.

booking. `https://www.booking.com/`. Accessed: 2022-04-17.

Booking.com. `https://www.booking.com/`. Accessed: 2022-04-17.

Fact checks archive. `https://www.snopes.com/fact-check/`. Accessed: 2022-04-17.

Factcheck – channel 4 news. `https://www.channel4.com/news/factcheck/`. Accessed: 2022-04-17.

fatabyyano. `https://fatabyyano.net`. Accessed: 2022-04-17.

Github - thiagovas/bs-detector-dataset. `https://github.com/thiagovas/bs-detector-dataset`. Accessed: 2022-04-17.

Machine learning and optimization. `https://cims.nyu.edu/`

~munoz/files/ml_optimization.pdf. Accessed: 2022-04-17.

misbar.

Oberlo. https://www.oberlo.com/. Accessed: 2022-04-17.

politifact. http://www.politifact.com/. Accessed: 2022-04-17.

tripadvisor. https://www.tripadvisor.com.eg/. Accessed: 2022-04-17.

Abu Hammad, A. S. (2013). An approach for detecting spam in arabic opinion reviews.

Ahmed, H., Traore, I., and Saad, S. (2018). Detecting opinion spams and fake news using text classification. *Security and Privacy*, 1(1):e9.

Al-Ahmad, B., Al-Zoubi, A., Abu Khurma, R., and Aljarah, I. (2021). An evolutionary fake news detection method for covid-19 pandemic information. *Symmetry*, 13(6):1091.

Al Zaatari, A., El Ballouli, R., ELbassouni, S., El-Hajj, W., Hajj, H., Shaban, K., Habash, N., and Yahya, E. (2016). Arabic corpora for credibility analysis. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 4396–4401.

Alazab, A., Alazab, M., Abawajy, J., Hobbs, M., et al. (2011a). Web application protection against sql injection attack. In *Proceedings of the 7th International Conference on Information Technology and Applications*, pages 1–7.

Alazab, M. (2020). Automated malware detection in mobile app stores based on robust feature generation. *Electronics*, 9(3):435.

Alazab, M., Alazab, M., Shalaginov, A., Mesleh, A., and Awajan, A. (2020a). Intelligent mobile malware detection using permission requests and api calls. *Future Generation Computer Systems*, 107:509–521.

Alazab, M., Alhyari, S., Awajan, A., and Abdallah, A. B. (2021). Blockchain technology in supply chain management: an empirical study of the factors affecting user adoption/acceptance. *Cluster Computing*, 24(1):83–101.

Alazab, M., Awajan, A., Mesleh, A., Abraham, A., Jatana, V., and Alhyari, S. (2020b). Covid-19 prediction and detection using deep learning. *International Journal of Computer Information Systems and Industrial Management Applications*, 12(June):168–181.

Alazab, M. and Batten, L. M. (2015). Survey in smartphone malware analysis techniques. *New threats and countermeasures in digital crime and cyber terrorism*, pages 105–130.

Alazab, M., Moonsamy, V., Batten, L., Lantz, P., and Tian, R. (2012). Analysis of malicious and benign android applications. In *2012 32nd International Conference on Distributed Computing Systems Workshops*, pages 608–616. IEEE.

Alazab, M., Venkatraman, S., Watters, P., Alazab, M., and Alazab, A. (2011b). Cybercrime: the case of obfuscated malware. In *Global security, safety and sustainability & e-Democracy*, pages 204–211. Springer.

Alazab, M. A. (2014). *Analysis on smartphone devices for detection and prevention of malware.* PhD thesis, Deakin University.

Alorini, D. and Rawat, D. B. (2018). Bayesian reasoning based malicious data discovery on gulf-dialectical arabic tweets. In *2018 IEEE International Symposium on Technology and Society (ISTAS)*, pages 133–138. IEEE.

Alorini, D. and Rawat, D. B. (2019). Automatic spam detection on gulf dialectical arabic tweets. In *2019 International Conference on Computing, Networking and Communications (ICNC)*, pages 448–452. IEEE.

Alzanin, S. M. and Azmi, A. M. (2019). Rumor detection in arabic tweets using semi-supervised and unsupervised expectation–maximization. *Knowledge-Based Systems*, 185:104945.

Anil Kumar, K., Anil, B., Rajath Kumar, U., Anand, C., and Aniruddha, S. (2018). Effective approaches for classification and rating of users reviews. In *Proceedings of international conference on cognition and recognition*, pages 1–9. Springer.

Barushka, A. and Hajek, P. (2019). Review spam detection using word embeddings and deep neural networks. In *IFIP International Conference on Artificial Intelligence Applications and Innovations*, pages 340–350. Springer.

Batten, L. M., Moonsamy, V., and Alazab, M. (2016). Smartphone applications, malware and data theft. In *Computational intelligence, cyber security and computational models*, pages 15–24. Springer.

Chapelle, O., Scholkopf, B., and Zien, A. (2009). Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews]. *IEEE Transactions on Neural Networks*, 20(3):542–542.

Conroy, N. K., Rubin, V. L., and Chen, Y. (2015). Automatic deception detection: Methods for finding fake news. *Proceedings of the association for information science and technology*, 52(1):1–4.

Dawson, M. (2015). A brief review of new threats and countermeasures in digital crime and cyber terrorism. *New Threats and Countermeasures in Digital Crime and Cyber Terrorism*, pages 1–7.

Elhadad, M. K., Li, K. F., and Gebali, F. (2020a). Detecting misleading information on covid-19. *Ieee Access*, 8:165201–165215.

Elhadad, M. K., Li, K. F., and Gebali, F. (2020b). An ensemble deep learning technique to detect covid-19 misleading information. In *International Conference on Network-Based Information Systems*, pages 163–175. Springer.

Elnagar, A., Khalifa, Y. S., and Einea, A. (2018). Hotel arabic-reviews dataset construction for sentiment analysis applications. In *Intelligent Natural Language Processing: Trends and Applications*, pages 35–52. Springer.

Faustini, P. H. A. and Covoes, T. F. (2020). Fake news detection in multiple platforms and languages. *Expert Systems with Applications*, 158:113503.

Ferreira, W. and Vlachos, A. (2016). Emergent: a novel data-set for stance classification. In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: Human language technologies*. ACL.

Geurts, P., Fillet, M., De Seny, D., Meuwis, M.-A., Malaise, M., Merville, M.-P., and Wehenkel, L. (2005). Proteomic mass spectra classification using decision tree based ensemble methods. *Bioinformatics*, 21(14):3138–3145.

Gräßer, F., Kallumadi, S., Malberg, H., and Zaunseder, S. (2018). Aspect-based sentiment analysis of drug reviews applying cross-domain and cross-data learning. In *Proceedings of the 2018 International Conference on Digital Health*, pages 121–125.

Guyon, I., Weston, J., Barnhill, S., and Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine learning*, 46(1):389–422.

Haouari, F., Hasanain, M., Suwaileh, R., and Elsayed, T. (2020). Arcov19-rumors: Arabic covid-19 twitter dataset for misinformation detection. *arXiv preprint arXiv:2010.08768*.

Hasan, M. K., Akhtaruzzaman, M., Kabir, S. R., Gadekallu, T. R., Islam, S., Magalingam, P., Hassan, R., Alazab, M., and Alazab, M. A. (2022). Evolution of industry and blockchain era: Monitoring price hike and corruption using biot for smart government and industry 4.0. *IEEE Transactions on Industrial Informatics*.

Hassan, R. and Islam, M. R. (2019). Detection of fake online reviews using semi-supervised and supervised learning. In *2019 International conference on electrical, computer and communication engineering (ECCE)*, pages 1–5. IEEE.

Hijazi, S., Kumar, R., Rowen, C., et al. (2015). Using convolutional neural networks for image recognition. *Cadence Design Systems Inc.: San Jose, CA, USA*, 9.

Jain, N., Kumar, A., Singh, S., Singh, C., and Tripathi, S. (2019). Deceptive reviews detection using deep learning techniques. In *International Conference on Applications of Natural Language to Information Systems*, pages 79–91. Springer.

Jardaneh, G., Abdelhaq, H., Buzz, M., and Johnson, D. (2019). Classifying arabic tweets based on credibility using content and user features. In *2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)*, pages 596–601. IEEE.

Li, J., Ott, M., and Cardie, C. (2013). Identifying manipulated offerings on review portals. In *Proceedings of the 2013 conference on empirical methods in natural language processing*, pages 1933–1942.

Lippmann, R. P. (1989). Pattern classification using neural networks. *IEEE communications magazine*, 27(11):47–50.

Maas, A., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., and Potts, C. (2011). Learning word vectors for sentiment analysis. In *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies*, pages 142–150.

Mani, S., Kumari, S., Jain, A., and Kumar, P. (2018). Spam review detection using ensemble machine learning. In *International Conference on Machine Learning and Data Mining in Pattern Recognition*, pages 198–209. Springer.

McCallum, A., Nigam, K., et al. (1998). A comparison of event models for naive bayes text classification. In *AAAI-98 workshop on learning for text categorization*, volume 752, pages 41–48. Citeseer.

Narayan, R., Rout, J. K., and Jena, S. K. (2018). Review spam detection using opinion mining. In *Progress in intelligent computing techniques: Theory, practice, and applications*, pages 273–279. Springer.

Oshikawa, R., Qian, J., and Wang, W. Y. (2018). A survey on natural language processing for fake news detection. *arXiv preprint arXiv:1811.00770*.

Ott, M., Cardie, C., and Hancock, J. (2012). Estimating the prevalence of deception in online review communities. In *Proceedings of the 21st international conference on World Wide Web*, pages 201–210.

Ott, M., Choi, Y., Cardie, C., and Hancock, J. T. (2011). Finding deceptive opinion spam by any stretch of the imagination. *arXiv preprint arXiv:1107.4557*.

Rayana, S. and Akoglu, L. (2015). Collective opinion spam detection: Bridging review networks and metadata. In *Proceedings of the 21th acm sigkdd international conference on knowledge discovery and data mining*, pages 985–994.

Sabbeh, S. F. and BAATWAH, S. Y. (2018). Arabic news credibility on twitter: An enhanced model using hybrid features. *journal of theoretical & applied information technology*, 96(8).

Saeed, R. M., Rady, S., and Gharib, T. F. (2022). An ensemble approach for spam detection in arabic opinion texts. *Journal of King Saud University-Computer and Information Sciences*, 34(1):1407–1416.

Salem, F. K. A., Al Feel, R., Elbassuoni, S., Jaber, M., and Farah, M. (2019). Fa-kes: A fake news dataset around the syrian war. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 13, pages 573–582.

Saumya, S. and Singh, J. P. (2018). Detection of spam reviews: a sentiment analysis approach. *Csi Transactions on ICT*, 6(2):137–148.

Shu, K., Sliva, A., Wang, S., Tang, J., and Liu, H. (2017). Fake news detection on social media: A data mining perspective. *ACM SIGKDD explorations newsletter*, 19(1):22–36.

Tandoc Jr, E. C. (2019). The facts of fake news: A research review. *Sociology Compass*, 13(9):e12724.

Thorne, J. and Vlachos, A. (2018). Automated fact checking: Task formulations, methods and future directions. *arXiv preprint arXiv:1806.07687*.

Vlachos, A. and Riedel, S. (2014). Fact checking: Task definition and dataset construction. In *Proceedings of the ACL 2014 workshop on language technologies and computational social science*, pages 18–22.

Wahsheh, H. A., Al-Kabi, M. N., and Alsmadi, I. M. (2013). Spar: A system to detect spam in arabic opinions. In *2013 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)*, pages 1–6. IEEE.

Wang, W. Y. (2017). " liar, liar pants on fire": A new benchmark dataset for fake news detection. *arXiv preprint arXiv:1705.00648*.

Xu, M., David, J. M., Kim, S. H., et al. (2018). The fourth industrial revolution: Opportunities and challenges. *International journal of financial research*, 9(2):90–95.