ISSN: 0067-2904

Detecting Fake News in Social Media: An Approach Utilizing Machine Learning to Uncover Disinformation

Mustafa Abdul-Razzaq Kareem *, Amer Abdulmajeed Abdulrahman
Department of Computer Science, College of Science, University of Baghdad, Baghdad, Iraq

Received: 11/3/2024 Accepted: 14/8/2024 Published: 30/7/2025

Abstract

The increasing number of untruths on social media has become a critical concern, affecting public sentiment and confidence. The broad spread of misleading information on the Internet and other social platforms presents a substantial barrier, exerting an impact on public sentiment, influencing political discussions, and eroding the reliability of information sources. Identifying false information on the X platform, previously known as Twitter, is an intricate task because of the network's attributes, such as conciseness, swift spread, and varied user engagements. Extracting crucial information from brief texts, such as tweets, is challenging, even with precise labeling. This study focuses on recognizing misinformation on social media platforms. The CIC Truth-Seeker Dataset 2023, one of the most extensive datasets in its category, contains over 134,000 labeled tweets. The study introduces novel methods in the field of short text classification, incorporating machine learning and natural language processing techniques (NLP). These techniques involve feature extraction using the term frequency-inverse document frequency (TF-IDF) algorithm after the dataset is preprocessed. The study then tests a number of machine learning models, including Random Forest RF, K-Nearest Neighbor KNN, Decision Tree DT, Logistic Regression LR, Naive Bayes NB, and stochastic gradient descent SGD, to see which ones can tell the most accurate difference between real and fake tweets. The findings demonstrated significant advancements in models designed to handle short text effectively, effectively addressing a practical issue such as automatically identifying fake content on social media platforms. Furthermore, we have achieved a significant advantage over previous research on the same dataset. When implementing the models on the news data, the random forest method attained the utmost accuracy at 93%, while the K-Nearest Neighbor strategy yielded a lower accuracy of 68%. This research paper aims to offer helpful information and practical answers to recognizing and reducing false news on social media platforms, specifically focusing on the X platform. Through a Truth-Seeker dataset, we will utilize machine learning methods to enhance previous text classification models.

Keywords: fake news detection, short texts classification, Machine Learning, Natural language processing, social media platforms, Random Forest, K-nearest neighbors, Logistic Regression, Stochastic Gradient Descent, Decision Tree, Naive Bayes.

اكتشاف الأخبار المزيفة في وسائل التواصل الاجتماعي: نهج يستعمل التعلم الآلي لكشف المعلومات المضللة

*Email: mostafa.abd2201m@sc.uobaghdad.edu.iq

مصطفى عبد الرزاق كريم*, عامر عبد المجيد عبد الرحمن

قسم الحاسوب، كلية العلوم، جامعة بغداد، بغداد، العراق

الخلاصة

أصبح العدد المتزايد من الأكاذيب على وسائل التواصل الاجتماعي مصدر قلق بالغ، مما يؤثر على مشاعر الجمهور وثقته. يمثل الانتشار الواسع للمعلومات المضللة على الإنترنت ومنصات التواصل الاجتماعي الأخرى حاجزًا كبيرًا، ويؤثر على المشاعر العامة، ويؤثر على المناقشات السياسية، ويؤدي إلى تآكل موثوقية مصادر المعلومات. يعد تحديد المعلومات الكاذبة على منصة X، المعروفة سابقًا باسم Twitter، مهمة معقدة بسبب سمات الشبكة، مثل الإيجاز والانتشار السريع وتفاعلات المستخدمين المتنوعة، ومن الصعب استخراج المعلومات المهمة من النصوص المختصرة مثل التغريدات، حتى مع التصنيف الدقيق. تركز هذه الدراسة على التعرف على المعلومات الخاطئة الموجودة على منصات التواصل الاجتماعي باستخدام واحدة من أكثر مجموعات البيانات شمولاً في فئتها، مع أكثر من 134000 تغريدة تحمل عنوان CIC Truth-Seeker Dataset 2023 تقدم الدراسة أساليب جديدة في مجال تصنيف النص القصير تتضمن التعلم الآلي وتقنيات معالجة اللغة الطبيعية، تتضمن هذه التقنيات استخراج الميزات باستعمال خوارزمية TF-IDF بعد المعالجة المسبقة لمجموعة البيانات، ثم تستعمل الدراسة عددًا من نماذج التعلم الآلي المختلفة (RF, KNN, LR, SGD, DT, NB) لتقييم قدرة النماذج على التمييز بين التغريدات الحقيقية والمزيفة بأعلى مستوى من الدقة. وأظهرت النتائج تطورات كبيرة في النماذج المصممة للتعامل مع النص القصير بفعالية، ومعالجة مشكلة عملية مثل التعرف تلقائيًا على المحتوى المزيف على منصات التواصل الاجتماعي. علاوة على ذلك، فقد حققنا ميزة كبيرة مقارنة بالأبحاث السابقة التي أجريت على نفس مجموعة البيانات، فعند تطبيق النماذج على البيانات الإخبارية، حققت طريقة RF أقصى دقة بنسبة 93%، في حين حققت استراتيجية KNN أقل دقة قدرها 68%. تهدف هذه الورقة البحثية إلى تقديم معلومات مفيدة وإجابات عملية للمسألة المعقدة المتمثلة في التعرف على الأخبار الكاذبة والحد منها على منصات التواصل الاجتماعي، مع التركيز بشكل خاص على منصة X من خلال Truth-Seeker Dataset ومن خلال استعمال أساليب التعلم الآلي لتحسين نماذج التصنيف السابقة.

1. Introduction

Spreading fake news has become a major obstacle on social media. This issue undermines public trust and information reliability. Information is abundant and fast, so traditional fact-checking systems cannot keep up. Therefore, contemporary technology is essential. Smartphones' popularity allows for anytime access, in contrast to traditional media.
Furthermore, they provide social contact with acquaintances, family, and even new people via comment threads, which contain remarks, arguments, and approval and disapproval buttons. Social media's dominance in news dissemination allows for the widespread sharing of fake news, as well as its use in political manipulation [1]. Nevertheless, by applying unique technologies and features, social media platforms can propagate misleading information on a large scale. Erroneous information can be intentional or unintentional. Depending on the originator's goal, rumors can be true or false. Fake news is misinformation, unlike rumors [2].

In contemporary times, social media has become a powerful tool for spreading misinformation. The user base has grown substantially, with an annual growth rate of 9.9 percent, resulting in an average of 13 fresh users per second. Until 2020, almost 50% of US adults attended events via social media, but in 2018, just 20% of them indicated a frequent dependence on digital platforms for news. The significance of social media has consistently increased over the past several years, and there is no indication of this trend slowing down. Social media is pervasive in America, Europe, and Asia, with over 4.5 billion people actively

using these platforms as of October 2021. Additionally, the majority of social media users tend to be younger. Nearly 90% of individuals aged 18 to 29 employed several kinds of social media [3], [4]; the global social media readership topped 4.59 billion in 2022.

This number is expected to approach six billion by 2027, suggesting extensive use worldwide. The global social media utilization rate was 59% in January 2023. Internet users spend 151 minutes every day on social media and messaging apps, up 40 minutes from 2015. Facebook was the first social network, with over a billion users and three billion monthly subscriptions. Meta Platforms owns and manages Facebook, Instagram, WhatsApp, and Facebook Messenger, with over one billion monthly active users. Figure 1 shows the most popular social media users [5]. The US-based social media network X, formerly Twitter, launched in 2006. Known as the Internet's SMS, this website ranked among the top ten most visited in 2013. Twitter users may not be as numerous as Facebook users. In the second quarter of 2021, 206 million people used Twitter every day. This statistic demonstrates platform users' active engagement and public dissemination of information and thoughts [4]. Tweets can have up to 280 characters and no age limit. Many X accounts are public, allowing unrestricted viewing of each other's content. About 23% of internet-connected adults use this social networking site every day [6]. Figure 2 shows over 368 million monthly active users worldwide. The estimated aggregate for 2024 is 335 million, down 5% from 2022 [7]. Untrue news and rumors on social media affect public opinion, personal choices, user trust, and public conversation. An automated method that detects disinformation on social media is needed to reduce its negative impact [8]. Fake news detection predicts deception in an article, narrative, or publication. The Natural Language Processing research community focuses on disinformation detection. NLP classifies news stories as true or false [9].

Computers classify text using algorithms [10]. To train themselves to recognize patterns of fake news, machine learning systems use misinformation indicators. These algorithms can help identify and report deceptive content early on, stopping its spread. They may also automate fact-checking by comparing news reports or social media posts to credible sources and data. This can help identify errors and prevent misinformation. This study uses the Truth-Seeker Dataset, which contains over 134,000 news tweets on the X platform, to test the capacity to recognize fake news. Tweet character limits may make it difficult for models to interpret context and meaning. Natural language processing may struggle to classify tweets. To address these challenges, this work will preprocess data to separate relevant information and improve NLP models. Next, we will manually classify tweets using machine learning to match human annotation accuracy. We evaluated the models' accuracy, precision, and recall. We benchmark different machine learning methods to evaluate their performance. The study begins with a literature review of current work on the topic. Next, use the Truth-Seeker dataset, pre-processing, feature extraction, and classification techniques. We explain the study's findings and experiments below.

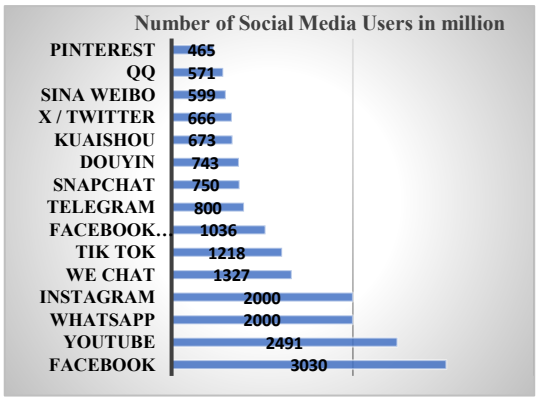

Figure 1: The user base of the largest social media sites [5]

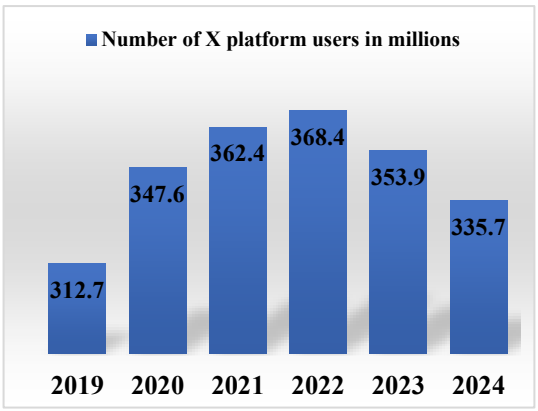

Figure 2: The users on the X platform [7].

2. Related work

Researchers have proposed many approaches to detect and expose false information and rumors on social media platforms. We'll explore some of these works below:

In [11], they established a binary machine learning challenge to identify deceptive news on the Twitter platform. The findings suggest that accepting certain inaccuracies in the labels makes it feasible to overcome the difficulty of obtaining large training datasets for false news classification while still attaining classifiers with excellent performance. The F1 score achieves 77% accuracy by analyzing only a single tweet. Considering the user's account details, the F1 score has the potential to reach a maximum of 90%. Manually annotating tweets as fake or actual news incurs significant costs and requires a substantial amount of time. Training samples are imprecise and full of irrelevant data.

Researchers in this study [12] utilized supervised classifiers to identify fraudulent information in Arabic tweets by employing machine-learning techniques on a publicly available Arabic dataset. To do this, they created several characteristics and classified them into two separate categories: content-based attributes and user-based qualities. They utilize four machine learning methods: DT, RF, AB, and LR throughout the learning process. The logistic regression model has a maximum recall rate of 83%. Conversely, the random forest method attains the maximum level of accuracy, scoring 76%. The model's accuracy considers the presence of noise and uncertainty in brief Arabic tweets.

In this study [13]. The authors gathered the information from online social forums like Facebook and Twitter. The compilation included news pieces from several domains to protect most news information rather than only classifying it as international news. The authors employed an artificial intelligence approach to develop fully automated news classifiers. Consequently, the hybrid SVM is the most efficient machine learning classification approach for identifying positive false news. The NLP model effectively utilized TF and TF-IDF, providing detailed descriptions. The suggested model achieved a maximum accuracy of 91.23% by employing unigram features and a hybrid SVM classifier. The study's tiny sample size may limit its applicability.

Researchers conducted a study on Persian tweets [14]. They generate a graph that illustrates the correlation between each user's followers and those of all users. They use the concept of information gain to assess content-driven and user-based characteristics, and they employ a variety of models for categorization. Two separate experiments examine the impact of two distinct sets of structural and content-based characteristics in identifying Persian rumors. The initial trial demonstrates a precision rate of approximately 70%, only relying on structural

features, while the second experiment achieves a precision rate above 80% by using both categories of data. Current natural language processing methods have limited the experiment's Persian potential.

In [15], the authors employed supervised machine-learning techniques to identify counterfeit news. The authors utilized three distinct real-world datasets for evaluation. This study introduces a biphasic methodology for identifying deceptive content on social media platforms. The first stage of the strategy involves using a large number of preprocessing procedures on the dataset. During the next stage, 23 supervised artificial intelligence algorithms were implemented. It was concluded that the decision tree algorithm exhibited superior performance compared to the other algorithms in terms of accuracy, precision, and F-measure. [16] suggested an innovative machine learning approach for identifying false news that enhances accuracy by up to 4.8% by integrating elements about news content and the social setting. The author implements his methodology on a Facebook Messenger chatbot and validates its efficacy through a practical application, attaining a detection accuracy of 81.7% in identifying fraudulent news.

The author of this study [17] concentrates on identifying false news and satire by introducing a system that combines machine learning and human input to detect potentially misleading content. They have chosen five distinct machine learning models, including LR, SVM, RF, neural networks, and gradient-boosting classifiers, to classify fake news and satire. The neural network model achieved a maximum accuracy of 81.64%, surpassing baseline values by 2.54%. Acknowledging the time-consuming and costly crowdsourcing process to achieve higher accuracy is essential. Furthermore, this study used a limited dataset; applying the model to a larger dataset could alter the outcomes.

The author in [18] proposed using the PHEME dataset, which comprises non-rumors and rumors about five significant events. Additionally, they developed an algorithm for identifying rumors in tweets. The analysis began with the evaluation and categorization of a variety of user and content characteristics. They employ natural language processing (NLP) techniques to generate specific qualities associated with the material. Subsequently, they employed diverse combinations of variables to train numerous ML models, including SVM, RF, and NB. Finally, they evaluated and contrasted the performance of the models. On one occasion, the models were 78% accurate. They suggest improving precision and data processing to integrate a fully automated rumor detection system into a microblogging platform.

The research [19]. Create a unique dataset from real-world sources, encompassing approximately 25,000 articles containing authentic and falsified information. Note that this dataset is considered small in size. The procedure entailed extracting linguistic characteristics, such as n-grams, from textual articles. We then trained many machine-learning models, including KNN, SVM, LR, LSVM, DT, and SGD. The SVM and logistic regression models achieved an accuracy of 92%.

The author [20] generated novel datasets named "Truth-Seeker," which encompass over 180,000 labels spanning from 2009 to 2022. These labels were assigned to tweets and categorized into two categorization schemes: one with five labels and another with three labels. The author employed Amazon Mechanical Turk for this classification process. The dataset was subjected to various validation stages to confirm its accuracy as a ground-truth standard. After that, the author came up with and tested a bunch of different machine learning and deep learning algorithms, such as different versions of BERT-based models and six different

machine learning models (DT, RF, KNN, BN, AB, and LR), to see how well they could tell the difference between real and fake tweets in both groups. The objective was to identify the versions that yielded the most favorable outcome metrics. The outcomes indicate that the Random Forest algorithm achieved the maximum accuracy rate of 70%, while the AdaBoost algorithm achieved the lowest accuracy rate of 59%. The BERTWEET model gets the highest accuracy, 96%, among deep learning models. However, the dataset's "unknown" category may misrepresent tweet classifications.

3. The Proposed Approach

The proposed method's primary architecture has three essential components: dataset, feature extraction, and classification, as depicted in Figure 2. The initial section pertains to the Truth-Seeker Dataset [20], which will be utilized in the trials. This dataset contains news articles from the X platform. The second stage is feature extraction, which employs two primary groups of features: linguistic and statistical. The field of linguistics encompasses several methods, such as tokenization of words, stop-word elimination, stemming, and normalization.

On the other hand, statistical methods rely on TF-IDF methods. The third section focuses on the classification task, which involves categorizing the Truth-Seeker dataset into positive and negative classes. Six classifiers have been utilized: Random Forest RF, K-Nearest Neighbors KNN, Logistic Regression LR, Stochastic Gradient Descent SGD, Decision Tree DT, and Naïve Bayes NB. The subsequent subsections will provide a more comprehensive analysis of each component [21].

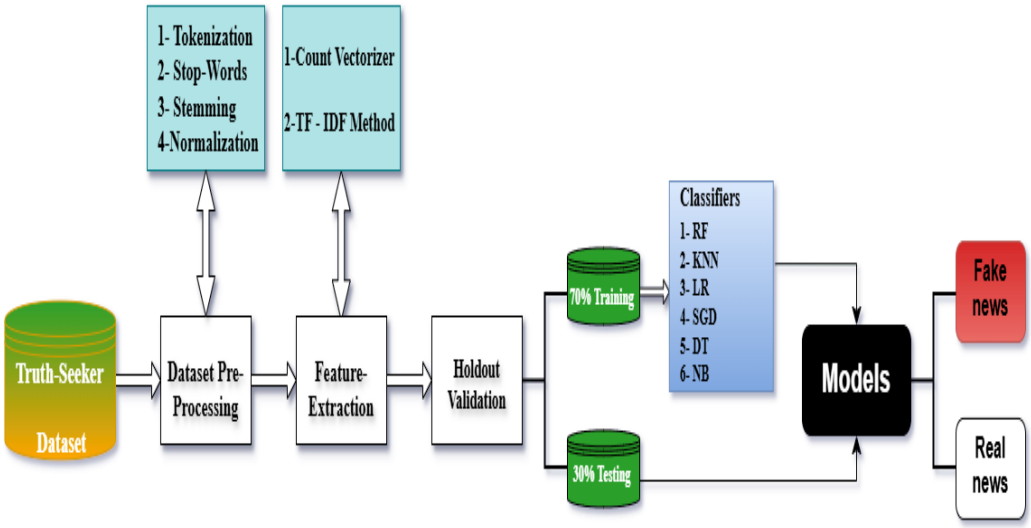

Figure 3: Proposed model.

4. Dataset

The dataset is the fundamental element for establishing the credibility and reliability of a machine-learning model. Nevertheless, the current fake news databases certainly have some constraints. Most current datasets require updates to accurately capture the sophisticated generation patterns employed by emerging false news authors. Furthermore, identifying numerous online social media users and messages as malevolent or dubious renders them inaccessible. Excelling on a dataset does not guarantee that any model is suitable for further data input [20]. Our study is mainly based on a thorough and inclusive dataset. To do this, we used the CIC Truth-Seeker Dataset 2023. The CIC Truth-Seeker Dataset 2023 is a benchmark dataset they created for analyzing true and false textual news information in social media posts. In Figure 3, it has over 134,000 English-language labeled tweets, making it one of the largest

datasets. The dataset was carefully vetted, utilizing a three-factor active learning verification process that included 456 unique and highly talented Amazon Mechanical Turks to label each tweet. In addition, the dataset contains three supplementary social media metrics: the bot score, credibility score, and impact score, to help analyze the trends and features of Twitter users. Notably, the PolitiFact dataset was used to create the Truth-Seeker dataset, which included tweets associated with both genuine and fraudulent news. They mostly used Amazon Mechanical Turk for crowdsourcing to determine the majority consensus on whether a tweet is actual or false news; this has resulted in the development of one of the most extensive ground truth datasets for identifying false news on the X platform [22].

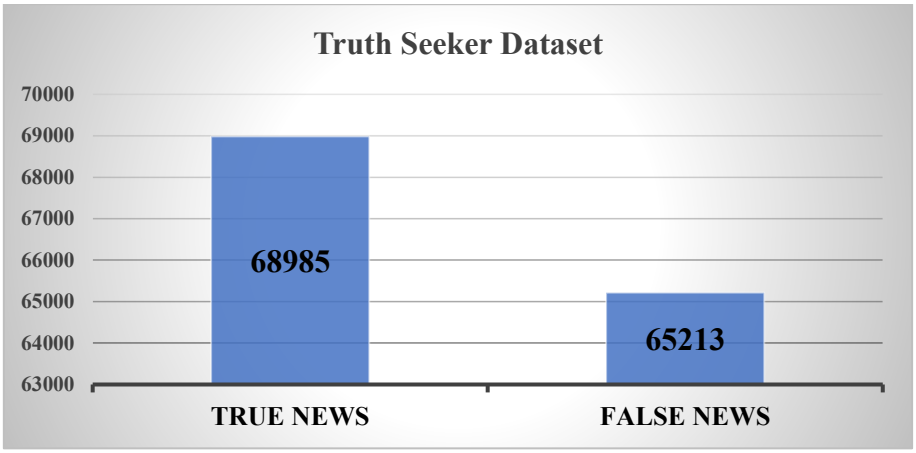

Figure 4: The Truth-Seeker data set.

5. Data Set Preprocessing

Data preparation refers to the methodologies and processes used to convert unprocessed data into a suitable format for processing or utilization in machine learning algorithms. The data preprocessing stage involves several actions to prepare the documents adequately. Text pre-processing is an essential stage in the classification process because it allows for the elimination of unnecessary words from documents, which often hinders the speed, accuracy, and efficiency of the classification process. The method of preprocessing documents used in this study consists of four steps: tokenization, elimination of stop words, stemming, and normalization. Those steps are crucial for building a text dataset that is clean, standardized, and meaningful. These methods will enhance the performance and interpretability of machine learning models when applied to text data [23].

Data Representation Overview

In the subsequent sections, we will detail the steps taken to represent the data effectively for our machine learning tasks. This paper provides a concise summary of the data transformation and representation, as well as the NLP tools and techniques used:

1. Data Loading and Cleaning: To ensure data quality, we load the dataset and remove any missing values.
2. Text Normalization: To maintain uniformity, this entails removing punctuation and numbers and converting text to lowercase.
3. Stop Word Removal: We eliminate common and domain-specific stop words to focus on more informative terms.
4. POS Tagging and NER: To apply part-of-speech tagging and named entity recognition to understand the grammatical structure and identify entities.

5. Steaming: This technique reduces words to their root forms, treating different forms of a word as a single item.
6. Vectorization: Text is converted into numerical format using Count Vectorizer and N-grams to capture word sequences.
7. TF-IDF Transformation: We apply TF-IDF to adjust word counts based on their frequency across the dataset.
8. Data Splitting: To evaluate model performance, we divide the dataset into training and testing sets.
Each of these steps is crucial for transforming raw text into a structured and meaningful format that enhances the efficiency, accuracy, and interpretability of our machine learning models. The following sections provide detailed explanations and implementations of these steps.

5.1 Tokenization

Tokenization divides text into separate tokens, which are individual words, making examining and altering the data more convenient. Machine learning models can then use significant characteristics extracted from the text. Removing stop words and applying stemming makes it easier to clean and preprocess tokens. This procedure ensures consistency in textual information by standardizing the method of dividing the text into individual words, thereby increasing the dependability of the following processing stages. Tokenization is an essential initial procedure in text pre-processing, which involves preparing unprocessed text for sophisticated analysis and modeling. The primary technique of tokenization is to divide a text into relevant chunks. Tokens are the designated labels for these individual components. An effective strategy involves segmenting a substantial chunk of the message into discrete words or phrases. Determining the principal function can categorize the incoming text into specific tokens based on our predetermined criteria [24].

5.2 Removal of Stop-Words

Every text file with words from a specific natural language has unique characteristics. All text files of this type have special characteristics involving the use of stop-words, as shown in Table 1. The first use of stop-word elimination dates back to Hans Peter Luhan in 1957. He proposed the classification of words in written natural language into two categories: keyword terms and non-keyword terms, commonly referred to as stop words. Stop-words are frequently occurring terms that lack meaningful or descriptive value compared to other words in a document. Stop words encompass prepositions, interjections, conjunctions, and numerals. The main goal of removing stop-words using pre-defined stop-word lists is to reduce disruption in textual data [25]. In addition, after removing stop words, Named Entity Recognition (NER), Part of Speech (POS) tagging, and N-grams maintain the meaning of the news. These methods ensure that removing stop words does not distort news.

The NLTK library's English stop words set includes common words.	Stop wards that removed from the news
{'a', 'about', 'above', 'after', 'again', 'against', 'all', 'am', 'an', 'and', 'any', 'are', "aren't", 'as', 'at', 'be', 'because', 'been', 'before', 'being', 'below', 'between', 'both', 'but', 'by', "can't", 'cannot', 'could', "couldn't", 'did', "didn't", 'do', 'does', "doesn't", 'doing', "don't", 'down', 'during', 'each', 'few', 'for', 'from', 'further', 'had', "hadn't", 'has', "hasn't", 'have', "haven't", 'having', 'he', "he'd", "he'll", "he's", 'her', 'here', "here's", 'hers', 'herself', 'him', 'himself', 'his', 'how', "how's", 'i', "i'd", "i'll", "i'm", "i've", 'if', 'in', 'into', 'is', "isn't", 'it', "it's", 'its', 'itself', "let's", 'me', 'more', 'most', "mustn't", 'my', 'myself', 'no', 'nor', 'not', 'of', 'off', 'on', 'once', 'only', 'or', 'other', 'ought', 'our', 'ours', 'ourselves', 'out', 'over', 'own', 'same', "shan't", 'she', "she'd", "she'll", "she's", 'should', "shouldn't", 'so', 'some', 'such', 'than', 'that', "that's", 'the', 'their', 'theirs', 'them', 'themselves', 'then', 'there', "there's", 'these', 'they', "they'd", "they'll", "they're", "they've", 'this', 'those', 'through', 'to', 'too', 'under', 'until', 'up', 'very', 'was', "wasn't", 'we', "we'd", "we'll", "we're", "we've", 'were', "weren't", 'what', "what's", 'when', "when's", 'where', "where's", 'which', 'while', 'who', "who's", 'whom', 'why', "why's", 'with', "won't", 'would', "wouldn't", 'you', "you'd", "you'll", "you're", "you've", 'your', 'yours', 'yourself', 'yourselves'}	['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you', "you're", "you've", "you'll", "you'd", 'your', 'yours', 'yourself', 'yourselves', 'he', 'him', 'his', 'himself', 'she', "she's", 'her', 'hers', 'herself', 'it', "it's", 'its', 'itself', 'they', 'them', 'their', 'theirs', 'themselves', 'what', 'which', 'who', 'whom', 'this', 'that', "that'll", 'these', 'those', 'am', 'is', 'are', 'was', 'were', 'be', 'been', 'being', 'have', 'has', 'had', 'having', 'do', 'does', 'did', 'doing', 'a', 'an', 'the', 'and', 'but', 'if', 'or', 'because', 'as', 'until', 'while', 'of', 'at', 'by', 'for', 'with', 'about', 'against', 'between', 'into', 'through', 'during', 'before', 'after', 'above', 'below', 'to', 'from', 'up', 'down', 'in', 'out', 'on', 'off', 'over', 'under', 'again', 'further', 'then', 'once', 'here', 'there', 'when', 'where', 'why', 'how', 'all', 'any', 'both', 'each', 'few', 'more', 'most', 'other', 'some', 'such', 'no', 'nor', 'not', 'only', 'own', 'same', 'so', 'than', 'too', 'very', 's', 't', 'can', 'will', 'just', 'don', "don't", 'should', "should've", 'now', 'd', 'll', 'm', 'o', 're', 've', 'y', 'a', 'are', "aren't", 'could', "couldn't", 'did', "didn't", 'does', "doesn't", 'had', "hadn't", 'has', "hasn't", 'have', "haven't", 'is', "isn't", 'ma', 'might', "mightn't", 'must', "mustn't", 'need', "needn't", 'sha', "shan't", 'should', "shouldn't", 'was', "wasn't", 'were', "weren't", 'won', "won't", 'would', "wouldn't"]

5.3 Stemming

Stemming is the process of converting words to their root form, resulting in a decrease in the number of word categories or types in the dataset. For instance, we will abbreviate the phrases "dancing," "dance," and "dancer" as "dance." Stemming enhances the efficiency and speed of classification [26]. This paper employed the Port Stemmer algorithm due to its high level of accuracy. Martin Porter created the Porter stemmer, a widely used stemming algorithm. It employs a sequence of regulations to eliminate suffixes from words. The goal is to create a relationship between related word forms, even if the result is not a valid word.

5.4 Normalization

Transforming a text into an official form includes removing any disruptions or irregularities, such as dates, whitespaces, abbreviations, acronyms, and diacritics, and generating standardized criteria for classification. Before conducting a meaningful analysis and constructing a solid classifier, it is imperative to preprocess the data by removing any irrelevant or noisy elements. This step includes converting all text to lowercase, removing symbols, handling URLs and mentions, and eliminating extra spaces. The goal of this stage in the

cleaning process is to enhance the quality of the data. Data cleansing is a crucial stage in significant NLP operations; improving the quality of textual data and ensuring the reliability of statistical analysis boosts overall effectiveness. The purpose of the method we employ for cleaning is to refine the data collection by filtering out irrelevant information and extracting just the relevant terms [27], [28].

6. Leveraging the TF-IDF Method for Feature Extraction

Feature extraction is a critical process in deceptive information classification. The procedure entails extracting distinctive features from the preprocessed textual input. A crucial text processing component is transforming a given text into a vector representation using geographical information. Various research papers have utilized feature-based classification techniques to improve the detection of fraudulent news articles. Examining textual characteristics quickly identifies deceptive information. This research employs the TF-IDF technique, a notable and vital strategy. The term frequency technique considers the frequency of the word's occurrence in all papers within the document collection. Inverse Document Frequency (IDF), another frequently used term, is calculated by dividing the logarithm of the total number of documents in the dataset by the number of documents containing a specific word. Lexical characteristics are predominantly employed in TF-IDF vectorization to summarize the number of unique words and the occurrence rate of each word. Lexical aspects refer to elements such as pronouns, verbs, hashtags, and punctuation marks. After obtaining the TF and IDF values, it becomes feasible to evaluate the viability of the TF-IDF approach, as illustrated in equations (1, 2, 3) [8], [29].

$$Tf(t,d) = \frac{term\ t\ count\ in\ d}{count\ of\ term\ T_d} \tag{1}$$

$$IDf(i) = log_2 \frac{N}{N_i} \tag{2}$$

$$TF - IDF = TF_i * (log_2 \left(\frac{N}{N_i}\right) \tag{3}$$

where TF = term frequency, t = term, and d = document.
N is the total number of papers in a document collection.
Ni is the frequency of the occurrence of the word I in a given group of documents.

7. Classification methods

Classification methods encompass a range of techniques and procedures employed to assign data into predicted classes or groups. In our work, we will specifically utilize supervised machine-learning methods for classification. After applying linguistic and statistical characteristics, we will train six supervised learning classifiers on 70% of the Twitter dataset. We will test the classifiers on the remaining 30%. Examine the classifier's ability to predict the class label, which can be positive or negative. We selected six classifiers: RF, KNN, LR, SGD, DT, and NB. With the 70% training dataset, the number of training examples increases. This method prevents misclassification and overfitting. Therefore, we can enhance the accuracy of data classification. Next, we'll explain each predictor separately [30].

7.1 Decision Tree

A decision tree is a predictive model commonly employed in machine learning. The main goal of a decision tree is to create a model that can precisely predict the values of target variables. The decision tree algorithm uses the variables generated during training to indicate

the target variables. A decision tree serves as a primary and straightforward method for classification. The decision tree algorithm utilizes elementary and direct concepts to address categorization difficulties. Typically, we construct a decision tree using a collection of attributes. Multiple varieties of decision trees exist, such as ID3 and C4.5. C4.5 is a very efficient classifier in data mining. C4.5 is a statistical classification algorithm. The C4.5 algorithm generates a decision tree by utilizing a provided set of training data. C4.5 uses a gain ratio and an information gain to prioritize potential tests. C4.5 is composed of four distinct steps [31]:

I. Select an attribute as the root.
II. Generate a branch for each value.
III. Place the dataset in the branch.
IV. Iterate the second process until all classes have equal values.

The formulas utilized in C4.5 are displayed here [31]:

$$Entropy\ (S) \sum_{i=1}^{n} -Pi * log_2\ Pi \tag{4}$$

S: Entropy, P: The proportion of classes in the output.

$$Gain(S, A) = Entropy\ (S) \sum_{i=1}^{n} \frac{|Si|}{|S|} * Entropy\ (S) \tag{5}$$

S: Compilation of instances.
A: Attribute of A case or problem.
|Si|: Denotes the numerical value of instances for I.
|S|: represents the number of cases in the set.

7.2 Random Forest

RF is a machine learning technique that belongs to the ensemble methods category. Each of the RF ensemble's classifiers is a decision tree classifier. Using the training dataset, the RF classifier constructs an ensemble of DTs. By aggregating the votes obtained from various decision trees, the system determines the ultimate label or class of the test object [32]. A solo tree classifier may be more accurate if it uses bootstrap aggregating and random data node selection when formed [33]. The steps of the RF classifier can be summarized as follows:

1. Sampling using the bootstrap method:

• Make lots of bootstrap samples D_i Train decision trees using the original dataset D.

$$D_i = \{(x_1, y_1), (x_2, y_2) \dots, (x_n, y_n)\} \tag{6}$$

2. Random Feature Selection:

• Consider a random subset of features F_i To divide each decision tree node i.

$$F_i = \{f_1, f_2, \dots, f_m\} \tag{7}$$

3. Training of Decision Trees:

• Utilize the features selected and text representations to train separate decision trees.
• Iteratively dividing nodes to remove impurities improves decision tree performance.

$$Impurity(D) = 1 - \sum_{k=1}^{c} p_k^2 \tag{8}$$

D represents the dataset at a node, C represents the number of classes, and pk represents the fraction of samples in class k.

4. Majority voting:

• Each decision tree in a given text source predicts a certain class.
• Majority vote across all decision trees determines the predicted class.

$$y\hat{} = mode(prediction\ generated\ by\ individual\ trees) \tag{9}$$

Where y^ represents the predicted class, the mode is the most frequently predicted class.

7.3 K-Nearest Neighbor

Another classifier, the KNN, uses the testing tweet's resemblance to earlier tweets from the training set. Alternatively, after training, we display tweets based on the TF-IDF of their keywords and then classify a given data point by considering the number of its closest neighbors. Each neighbor casts a vote for a class, and the prediction is assigned to the one with the most votes. In other words, a point's classification is determined by the majority consensus of its neighboring points [34]. KNN distance computations predominantly employ the Euclidean distance, whose formula is as follows:

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i + y_i)^2} \tag{10}$$

Where d is the Euclidean distance, X denotes a distinct datum extracted from the dataset. n denotes the overall quantity of dimensions, and Y denotes a predicted data point.

7.4 Logistic Regression

LR is a commonly used classification technique that addresses binary classification problems. The model is a statistical approach that utilizes a vector of variables to determine the magnitude of each variable's influence. We then use the weight to predict the type of fabricated information, represented as a word vector. Only when the dependent variable has two possible outcomes, also referred to as dichotomous or binary variables, can we apply LR. In linear regression LR, there is no explicit correlation between the dependent and independent variables, and the independent variable does not need to follow a normal distribution or have equal variation within a group [35]. We can mathematically express the LR hypothesis function in the following way [36]:

$$h_\theta(X) = \frac{1}{1+e^{-(\beta_0+\beta_1 X)}} \tag{11}$$

Below is an analysis of the terms:

- h(x): This denotes the probability that the dependent variable X assumes the value of 1.
- e: The natural logarithm fundamental constant, about 2.71828.
- $\beta_0,\beta_1\ldots,\beta_n$: represent the coefficients of the logistic regression model. These coefficients are calculated throughout the process of training the model.
- $X_1,X_2\ldots,X_n$: the autonomous variable or characteristic linked to the observation under consideration.

LR utilizes a sigmoid function to transform the output into a probability value. The objective is to minimize the cost function to achieve an optimal probability. The cost function is calculated using the subsequent equation [36]:

$$Cost\ (h_\theta(x), y) = \begin{cases} log(h_\theta(x)) & y = 1 \\ -log\big(1 - h_\theta(x)\big), & y = 0 \end{cases} \tag{12}$$

where: y is the actual label 0 or 1.

7.5 Stochastic Gradient Descent (SGD)

SGD is a kind of Gradient Descent (GD) that highlights random chance. Widely used, SGD has proven its superior performance in several machine learning tasks. The learning rate parameter determines the size of the next step to take when advancing in the gradient direction. We select just one sample to train the model during each iteration. Each iteration significantly reduces the time required to calculate the cost function (c^f) for a single training sample x^i, resulting in faster convergence towards the local minimum. It does this by adjusting the

parameters of the model for each input. x^i and its corresponding target class y^i[37], [38]. In SGD, the formula for adjusting the parameters (weights) is:

$$\theta_{i+1} = \theta_i - \alpha \cdot \nabla J\left(\theta_i;\ x^{(i)}, y^{(i)}\right) \quad (13)$$

where: θ_i Denotes the model parameters at the i-th iteration.
The learning rate, denoted by α, specifies the magnitude of the steps taken in the parameter space throughout the learning process.
The loss function, denoted as $J(\theta_i; x^{(i)}, y^{(i)}$ quantifies the disparity between the expected output and the true goal for the training example $(x^{(i)}, y^{(i)})$
The gradient of the loss function for the parameters θ_i, denoted as $\nabla J(\theta_i;\ x^{(i)}, y^{(i)})$, represents the direction and amount of the sharpest rise in the loss.

7.6 Naïve Bayes

The NB classifier utilizes the Bayes theorem and is a probabilistic classifier. Conditional probabilities with independent assumptions about its characteristics form the foundation of NBC. The Naïve Bayes classifier classifies data according to values found in the test data and training set, assuming categorical class labels. Applications for Naïve Bayes include spam filtering, text classification, and hybrid recommender systems with machine learning text categorization, which is the most effective method. The Naïve Bayes classifier is defined mathematically as follows:

$$P(X|E_1, \dots, E_n) = \frac{P(E_1, \dots, E_n|X)P(X)}{P(E_1, \dots, E_n)} \quad (14)$$

where E is the available evidence, and X is the likelihood of an event.
$P\ (E_1 \dots E_n\ |X)$ = probability, $P(X)$ equals Prior, $P\ (E_1 \dots E_n)$ = constant of normalization [39].

8. Performance Evaluation for Models

In machine learning and statistics, confusion matrices and evaluation metrics generally fall under the category of performance or model evaluation. These methods evaluate the efficacy of a predictive model, usually a classification model, by contrasting its predictions with the actual results. In machine learning and data analysis, confusion matrices and evaluation metrics are integral to model evaluation. Assessment measures and analysis are covered in the following subsections [40].

8.1 Confusion Matrix

The confusion matrix, or mistake matrix, is a tabular representation that showcases a classification model's effectiveness and evaluates its performance. This scenario typically incorporates supervised and unsupervised learning, sometimes called a matching matrix. Since there are two classes, each feature in the matrix represents instances in an observed class, and each entry indicates cases in a predicted class or vice versa [41].

Table 1: Confusion Matrix.

	Predicted Positive Class	Predicted Negative Class
Actual Positive Class	True positive TP	False negative FN
Actual Negative Class	False positive FP	True negative TN

The confusion matrix displays the expected class in the row and the actual class in the column. In this case, we have classified accurate news as positive and fake news as unfavorable. Therefore, true positive refers to news that is true and correctly predicted to be accurate. False positive, on the other hand, denotes the mistaken identification of factually incorrect news as

valid. True negatives are cases in which news is correctly predicted to be unfavorable. Misinterpretation of real news as bad leads to false negatives [42].

8.2 Assessment Techniques

Various evaluation methods are often used to gauge the effectiveness of classifiers. The selection of a metric is contingent upon the classification of the task's characteristics and the precise objectives you aim to accomplish. Below are frequent classifier assessments:

8.2.1 Precision: The phrase for this is positive predictive value, which refers to the ratio of correctly detected positive tweets to the entire number of positive predictions. Equation determining the precision [43]:

$$Precision = \frac{TP}{TP+FP} \quad (15)$$

8.2.2 Recall: sensitivity, or actual positive rate, is a metric used in classifier analysis to accurately estimate the model's ability to detect all positive instances in the dataset. It formerly denoted the system's capacity to categorize inputs [44] accurately. The computation is executed via the given formula:

$$Recall = \frac{TP}{TP+FN} \quad (16)$$

8.2.3 F1 score: also known as the balanced F-score or F-measure, The harmonic average of the Recall and precision. We use it to ascertain each classifier's final performance metric value [45]. The values might range from 1, indicating the best, to 0, indicating the worst [46]. It imposes penalties on classifiers that excessively prioritize one attribute to the detriment of the other. This characteristic makes it a relevant indicator in circumstances where it is important to minimize false positives and false negatives, providing a thorough assessment of a classifier's performance. The formula for calculating the F1 score:

$$F1 = 2 \cdot \frac{Precision \times Recall}{Precision+Recall} \quad (17)$$

8.2.4 Accuracy: The assessment metric most frequently used in practice, whether for binary or multi-class classification tasks, is. We define classification accuracy as the ratio of accurately classified occurrences of false and real tweets to the total number of incorrectly and adequately classified cases [47]. The calculation formula:

$$Accuarcy = \frac{TP+TN}{TP+TN+FP+FN} \quad (18)$$

9. Result and Discussion

This study introduces a classification methodology to identify false news on social media platforms. We investigated using the Truth-Seeker dataset, which includes over 134,000 labels from 2009 to 2022, focusing on tweets. The dataset has two separate labels: true and false. We divided the dataset into a training set, which included 70% of the data, and a testing set, which included the remaining 30%. We used natural language processing techniques to analyze the data and identify the significant aspects. As mentioned, this study aims to evaluate the accuracy attained using the TF-IDF technique with various machine learning classifiers and then compare their performance. The paper presents and expresses the results of this study using six classifiers. Table 2 lists these findings and displays the outcomes of the six classifiers for the testing set using the metrics of precision, F1-score, Recall, and Accuracy. The Random Forest classifier achieves the highest accuracy of 93%. The Logistic Regression and Stochastic Gradient Descent classifiers also perform well, with accuracies of 92% and 91%, respectively, which are close to that of the Random Forest. The other three classifiers, namely Naive Bayes (NB), Decision Tree (DT), and K-Nearest Neighbors (KNN), attained accuracy rates of 90%, 89%, and 72%, respectively. The Random Forest, KNN, and Decision Tree models exhibit a high level of accuracy throughout the training phase. The model exhibits exceptional performance on both the training data and unseen data, with minimal disparity. Compared to tree-based models and KNN, Naive Bayes, Logistic Regression, and SGD have a smaller difference between their training and testing accuracies. This means they are less likely to

overfit. Table 3 presents the outcomes of the six classifiers for the training set. Table 4 and Table 5 display the confusion matrix results for the models used on the training and testing sets, respectively.

Compared to earlier research, Figure 5 shows our methodology's outstanding performance in recognizing bogus news. Several classifiers were evaluated: The decision tree model's accuracy improved from 62% to 88%. The Random Forest model outperformed the prior studies' 70% accuracy with 93%. The K-Nearest Neighbors (KNN) algorithm improved from 62% to 66%. Logistic regression performed well, with an accuracy of 92%, up from 63% in earlier experiments. SGD has 91% accuracy, while Naive Bayes has 89%. These data demonstrate our technique's ability to detect fake news using multiple classifier models. The large precision improvements demonstrate our approach's potential for real-world use in social media monitoring and disinformation detection.

Table 2: The outcomes of six classifiers for Testing set.

Classifier Name	Precision	F1-score	Recall	Accuracy
RF	0.93	0.93	0.93	0.93
LR	0.92	0.92	0.92	0.92
SGD	0.92	0.92	0.92	0.91
NP	0.9	0.9	0.9	0.89
DT	0.89	0.89	0.89	0.88
KNN	0.79	0.68	0.72	0.66

Table 3: The outcomes of six classifiers for Training set.

Classifier Name	Precision	F1-score	Recall	Accuracy
RF	100	100	100	0.99
LR	0.93	0.93	0.93	0.93
SGD	0.93	0.92	0.92	0.92
NP	0.91	0.9	0.9	0.9
DT	100	100	100	0.99
KNN	0.87	0.68	0.72	0.68

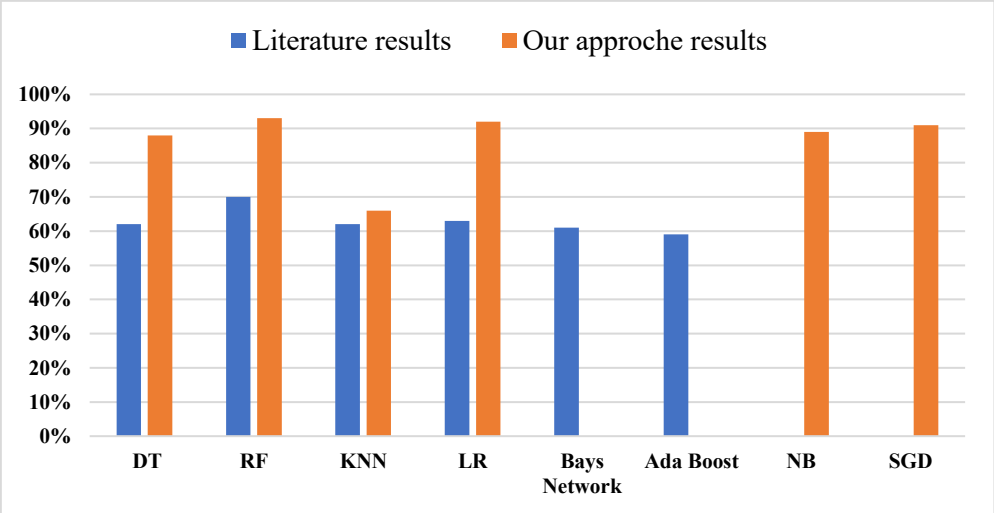

Figure 5: Comparative with prior research.

Table 4: Confusion matrix for Train set.

classifier	Actual \ Predicted	True	Fake
Random Forest	True	45658	70
	Fake	42	48168
K-Nearest Neighbor	True	17159	28569
	Fake	622	47588
Decision Tree	True	45686	42
	Fake	65	48145
Logistic Regression	True	42451	3277
	Fake	3072	45138
Naive Bayes	True	40805	4923
	Fake	3777	44433
Stochastic Gradient Descent	True	41879	3849
	Fake	3388	44822

Table 5: Confusion matrix for Testing set.

classifier	Actual \ Predicted	True	Fake
Random Forest	True	18012	1473
	Fake	1341	19434
K-Nearest Neighbor	True	6171	13314
	Fake	277	20498
Decision Tree	True	17182	2303
	Fake	2288	18487
Logistic Regression	True	17854	1631
	Fake	1507	19268
Naive Bayes	True	17152	2333
	Fake	1872	18903
Stochastic Gradient Descent	True	17717	1768
	Fake	1548	19227

10.Case Study: Disinformation and Fake News in the U.S. 2020 Presidential Election[48].

Overview

The proliferation of false information is indisputable in light of the emergence of social media. Disinformation, which includes fake news, refers to intentional misinformation that aims to intentionally mislead, deceive, or confuse individuals. False information garnered significant attention during the 2016 United States presidential election, and we anticipated a further increase in its prevalence in the 2020 presidential election.

Goals

This case study seeks to address two fundamental inquiries:

What type of disinformation or fake news did the United States 2020 presidential election employ?

How did the dissemination of these assertions affect the election results?

Methodology

Data collection

Fact-checking data was gathered between July 1st and September 25th, 2020, utilizing the reputable factcheck.org website. A grand total of 327 counterfeit news stories were found. A total of 239 articles were specifically focused on the U.S. 2020 presidential election, with keywords such as "Trump," "Biden," "election," and "vote."

Data analysis

The RapidMiner software was utilized to perform topic modeling utilizing Latent Semantic Analysis (LSA). The procedure consisted of multiple processes, namely tokenization, removal of stop words, stemming, and transformation using term frequency-inverse document frequency (TF-IDF).

Discoveries

The investigation revealed four primary topics regarding the 2020 United States presidential election:

COVID-19: The dissemination of false or inaccurate information about the pandemic and its impact on elections.

Trump: Refers to assertions pertaining to Donald Trump, encompassing his conduct and governmental measures.

Biden: Pertaining to assertions concerning Joe Biden, encompassing his tax proposals and declarations.

Voting Process: The dissemination of incorrect information about the voting process, such as mail-in ballots and election fraud.

Conversation

The study revealed that disinformation and fake news had a substantial influence on popular perception during the election. Disinformation disseminates more rapidly and extensively than accurate information, especially when it pertains to political matters. This emphasizes the importance of distinguishing genuine information from fabricated narratives in order to make well-informed decisions.

In conclusion

The case study emphasizes the critical role of precise information in political events and elections. The results underscore the necessity for strong strategies to counteract disinformation and enlighten the public on how to identify and dismiss false information.

10. Conclusion

This study investigates the utilization of diverse natural language processing (NLP) approaches and machine learning algorithms for the categorization of text data. We undertook a sequence of crucial procedures, including tokenization, stop word removal, and stemming, to purify and uniformize the dataset of various forms of text for subsequent analysis. We employed two primary methods for feature extraction. We used the Count Vectorizer and Term Frequency-Inverse Document Frequency (TF-IDF). The count vectorizer converted the text data into a sparse matrix that represents the frequency of each word in the corpus. This strategy is especially valuable for displaying textual data in a clear and easily understandable manner. We then utilized the TF-IDF transformation to modify these frequencies, highlighting terms

that are more significant in specific documents than the entire collection of documents. TF-IDF's use on brief news items has demonstrated its advantages by accurately assigning weights to phrases and improving the model's performance. TF-IDF effectively detected and highlighted important terms in the short documents, resulting in improved classification accuracy and more dependable predictions. We employed six distinct machine learning algorithms: random forest, logistic regression, K-nearest neighbors (KNN), decision tree, multinomial Naive Bayes, and stochastic gradient descent (SGD). The goal of utilizing these six strategies was to accurately assess their efficacy and determine the most appropriate model for our text classification challenge. Every algorithm possesses distinct advantages and processes the data in a unique manner, resulting in a wide range of performance measures for evaluation. Through the analysis of these diverse methodologies, we can ascertain the most resilient and dependable model for our particular dataset and classification objectives. The results demonstrated various degrees of accuracy and performance among the different models, with certain models surpassing others in terms of precision, recall, and F1-score. The random forest and logistic regression models exhibited promising outcomes, showcasing their proficiency in efficiently managing the text classification problem. To summarize, this study emphasizes the importance of meticulous text preprocessing as well as the benefits of exploring various feature extraction techniques and machine learning algorithms to achieve the best possible outcomes in text classification assignments. The knowledge gained from this study can serve as a foundation for future research and advancement in the fields of NLP and text mining, potentially leading to the development of more sophisticated and precise text classification systems.

Future research will focus on curating collections of photos, videos, and text from X and other social media platforms. This large dataset will help us identify and combat fake news, photographs, and altered films. We use computer vision and natural language processing to build robust solutions that can ensure media accuracy and reliability on digital platforms.

References

[1] A. M. U. D. Khanday, Q. R. Khan, and S. T. Rabani, "Detecting textual propaganda using machine learning techniques," *Baghdad Science Journal*, vol. 18, no. 1, pp. 199–209, Mar.2021, doi: 10.21123/bsj.2021.18.1.0199.

[2] S. V. Balshetwar, R. S. Abilash, and R. Dani Jermisha, "Fake news detection in social media based on sentiment analysis using classifier techniques," *Multimedia Tools and Applications*, vol. 82, no. 23, pp. 35781–35811, Sep. 2023, doi: 10.1007/s11042-023-14883-3.

[3] E. Aïmeur, S. Amri, and G. Brassard, "Fake news, disinformation and misinformation in social media: a review," *Social Network Analysis and Mining*, vol. 13, no. 1, Feb. 2023, doi: 10.1007/s13278-023-01028-5.

[4] Y. Koul, K. Mamgain, and A. Gupta, "Lifetime of tweets: a statistical analysis," *Social Network Analysis and Mining*, vol. 12, no. 1, Aug. 2022, doi: 10.1007/s13278-022-00926-4.

[5] Stacy Jo Dixon, "Most popular social networks worldwide as of October 2023, ranked by number of monthly active users," *https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/.*

[6] S. T. Rabani, Q. R. Khan, and A. M. Ud Din Khanday, "Detection of suicidal ideation on Twitter using machine learning & ensemble approaches," *Baghdad Science Journal*, vol. 17, no. 4, pp. 1328–1339, Dec. 2020, doi: 10.21123/bsj.2020.17.4.1328.

[7] https://www.statista.com/statistics/303681/twitter-users-worldwide/, *"Number of X (formerly Twitter) users worldwide from 2019 to 2024."*

[8] S. Mishra, P. Shukla, and R. Agarwal, "Analyzing Machine Learning Enabled Fake News Detection Techniques for Diversified Datasets," *Wireless Communications and Mobile Computing*, vol. 2022, pp. 1-18, Hindawi Limited, Mar.2022. doi: 10.1155/2022/1575365.

[9] T. A. Wotaifi and B. N. Dhannoon, "An Effective Hybrid Deep Neural Network for Arabic Fake News Detection," *Baghdad Science Journal*, vol. 20, no. 4, pp. 1392–1401, Aug.2023, doi: 10.21123/bsj.2023.7427.

[10] F. A. Abdulghani and N. A. Z. Abdullah, "A Survey on Arabic Text Classification Using Deep and Machine Learning Algorithms," *Iraqi Journal of Science*, vol. 63, no. 1, pp. 409–419, Jan. 2022, doi: 10.24996/ijs.2022.63.1.37.

[11] S. Helmstetter and H. Paulheim, "Weakly supervised learning for fake news detection on Twitter," *in Proceedings of the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2018, Institute of Electrical and Electronics Engineers Inc*., pp. 274–277, 25. Oct. 2018, doi: 10.1109/ASONAM.2018.8508520.

[12] G. Jardaneh, H. Abdelhaq, M. Buzz and D. Johnson, "Classifying Arabic Tweets Based on Credibility Using Content and User Features," *2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)*, Amman, Jordan, pp. 596-601, 20. May. 2019, doi: 10.1109/JEEIT.2019.8717386.

[13] R. Setiawan et al., "Certain Investigation of Fake News Detection from Facebook and Twitter Using Artificial Intelligence Approach," *Wireless Personal Communications*, vol. 127, no. 2, pp. 1737–1762, Nov. 2022, doi: 10.1007/s11277-021-08720-9.

[14] S. Zamani, M. Asadpour, and D. Moazzami, "Rumor detection for Persian Tweets," *in 2017 25th Iranian Conference on Electrical Engineering, ICEE 2017, Institute of Electrical and Electronics Engineers Inc*., pp. 1532–1536, 20. Jul. 2017, doi: 10.1109/IranianCEE.2017.7985287.

[15] F. A. Ozbay and B. Alatas, "Fake news detection within online social media using supervised artificial intelligence algorithms," *Physica A: Statistical Mechanics and its Applications*, vol. 540, pp.123174, Feb. 2020, doi: 10.1016/j.physa.2019.123174.

[16] M. L. Della Vedova, E. Tacchini, S. Moret, G. Ballarin, M. DiPierro, and L. de Alfaro, "Automatic Online Fake News Detection Combining Content and Social Signals," *2018 22nd Conference of Open Innovations Association (FRUCT)*, Jyvaskyla, Finland, pp. 272-279, 20. Sep. 2018 doi: 10.23919/FRUCT.2018.8468301.

[17] S. Shabani and M. Sokhn, "Hybrid machine-crowd approach for fake news detection," *in Proceedings - 4th IEEE International Conference on Collaboration and Internet Computing, CIC 2018, Institute of Electrical and Electronics Engineers Inc*., pp. 299–306, 18. Nov. 2018, doi: 10.1109/CIC.2018.00048.

[18] A. Vijeev, A. Mahapatra, A. Shyamkrishna and S. Murthy, "A Hybrid Approach to Rumour Detection in Microblogging Platforms," *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Bangalore, India*, pp. 337-342, 02. Dec.2018, doi: 10.1109/ICACCI.2018.8554371.

[19] H. Ahmed, I. Traore, and S. Saad, "Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques," *Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments: First International Conference, ISDDC 2017, Vancouver, BC, Canada, October 26-28, 2017, Proceedings 1. Springer International Publishing,2017*, Vol 10618, pp. 127–138, 11 Oct. 2017, doi: 10.1007/978-3-319-69155-8_9.

[20] S. Dadkhah, X. Zhang, A. G. Weismann, A. Firouzi and A. A. Ghorbani, "The Largest Social Media Ground-Truth Dataset for Real/Fake Content: TruthSeeker," *in IEEE Transactions on Computational Social Systems*, pp. 1–15, 12 May. 2017, doi: 10.1109/TCSS.2023.3322303.

[21] A. S. A. AL-Jumaili and H. K. Tayyeh, "A hybrid method of linguistic and statistical features for Arabic sentiment analysis," *Baghdad Science Journal*, vol. 17, no. 1, pp. 385–390, Mar. 2020, doi: 10.21123/BSJ.2020.17.1(SUPPL.).0385.

[22] Y. Sharrab, D. Al-Fraihat, and M. Alsmirat, "Deep Neural Networks in Social Media Forensics: Unveiling Suspicious Patterns and Advancing Investigations on Twitter," *2023 3rd Intelligent Cybersecurity Conference (ICSC), San Antonio, TX, USA, 2023*, pp. 95-102, 18. Dec. 2023 doi: 10.1109/ICSC60084.2023.10349985.

[23] A. Abdulrahman and M. Baykara, "Fake News Detection Using Machine Learning and Deep Learning Algorithms," *2020 International Conference on Advanced Science and Engineering (ICOASE), Duhok, Iraq*, pp, 31. May. 2020, 18-23, doi: 10.1109/ICOASE51841.2020.9436605.

[24] A. Deshmukh, G. Mahto, J. Sharma, and H. Dedhia, "Multi-Label Classification of Fake News on Social Media," *International Journal of Innovative Science and Research Technology (IJISRT)*.,

vol. 7, no. 4, pp. 852-856, Apr. 2022. [Online]. Available: www.ijisrt.com. [Accessed: Apr. 24, 2024]. DOI: 10.5281/zenodo.6539310.

[25] S. S. Abdul-Jabbar and L. E. George, "Fast Text Analysis Using Symbol Enumeration and Hashing Methodology," *Iraqi Journal of Science*., vol. 58, no. 1B, pp. 345–354, Jan. 2022. [Online]. Available: https://ijs.uobaghdad.edu.iq/index.php/eijs/article/view/6169. [Accessed: Apr. 24, 2024].

[26] A. Hansrajh, T. T. Adeliyi, and J. Wing, "Detection of Online Fake News Using Blending Ensemble Learning," *Scientific Programming*, vol. 2021, Article ID 3434458, pp. 1-10, 2021. [Online]. Available: https://doi.org/10.1155/2021/3434458. [Accessed: Apr. 24, 2024].

[27] M. Q. Saadi and B. N. Dhannoon "Arabic Cyberbullying Detection Using Support Vector Machine with Cuckoo Search," *Iraqi Journal of Science*, vol. 64, no. 10, pp. 5322–5330, Oct. 2023, doi: 10.24996/ijs.2023.64.10.37.

[28] M. Alkhair, K. Meftouh, K. Smaïli, and N. Othman, "An Arabic Corpus of Fake News: Collection, Analysis and Classification," *in Arabic Language Processing: From Theory to Practice, ICALP 2019*, vol. 1108, K. Smaïli, Ed. Springer, Cham, 11. Oct. 2019, pp. 253-263. doi: 10.1007/978-3-030-32959-4_21.

[29] M. Zayno and A. M. Radhi, "Data Mining Methods for Extracting Rumors Using Social Analysis Tools," *Iraqi Journal of Science*, vol. 63, no. 8, pp. 3618–3627, Aug. 2022, doi: 10.24996/ijs.2022.63.8.36.

[30] N. S. M. Nafis and S. Awang, "The evaluation of accuracy performance in an enhanced embedded feature selection for unstructured text classification," *Iraqi Journal of Science*, vol. 61, no. 12, pp. 3397–3407, Dec. 2020, doi: 10.24996/ijs.2020.61.12.28.

[31] A. H. Mohammad, "Arabic Text Classification: A Review," *Modern Applied Science*, vol. 13, no. 5, p. 88, Apr. 2019, doi: 10.5539/mas. v13n5p88.

[32] Ankit and N. Saleena, "An Ensemble Classification System for Twitter Sentiment Analysis," *International Conference on Computational Intelligence and Data Science, in Procedia Computer Science, Elsevier BV*, vol.132, pp. 937–946, 8. Jun. 2018, doi: 10.1016/j.procs.2018.05.109.

[33] R. C. Chen, C. Dewi, S. W. Huang, and R. E. Caraka, "Selecting critical features for data classification based on machine learning methods," *Journal of Big Data*, vol. 7, no. 1, p. 52, 23 Jul 2020, doi: 10.1186/s40537-020-00327-4.

[34] L. A. Qadi, H. E. Rifai, S. Obaid, and A. Elnagar, "Arabic Text Classification of News Articles Using Classical Supervised Classifiers," *in 2019 2nd International Conference on New Trends in Computing Sciences (ICTCS), Amman, Jordan*, pp. 1-6, 05. Dec.2019 doi: 10.1109/ICTCS.2019.8923073.

[35] W. H. Bangyal, R. Qasim, N. u. Rehman, Z. Ahmad, H. Dar, L. Rukhsar, Z. Aman, J. Ahmad, "Detection of Fake News Text Classification on COVID-19 Using Deep Learning Approaches," *Computational and Mathematical Methods in Medicine*, vol. 2021, pp. 1-14, 15. Nov. 2021, https://doi.org/10.1155/2021/5514220.

[36] I. Ahmad, M. Yousaf, S. Yousaf, and M. O. Ahmad, "Fake News Detection Using Machine Learning Ensemble Methods," *Complexity*, vol. 2020, 17. Oct. 2020, doi: 10.1155/2020/8885861.

[37] B. Gaye, D. Zhang, and A. Wulamu, "Sentiment classification for employees reviews using regression vector-stochastic gradient descent classifier (RV-SGDC)," *PeerJ Computer Science*, vol. 7, p. e712, 23. Sep. 2021. https://doi.org/10.7717/peerj-cs.712

[38] A. A. Abdulrahman and M. K. Ibrahem, "Intrusion detection system using data stream classification," *Iraqi Journal of Science*, vol. 62, no. 1, pp. 319–328, Jan. 2021, doi: 10.24996/ijs.2021.62.1.30.

[39] A. Prabhat and V. Khullar, "Sentiment classification on big data using Naïve Bayes and logistic regression," *2017 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India*, pp. 1-5, 23. Nov. 2017, doi: 10.1109/ICCCI.2017.8117734.

[40] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means algorithm: A comprehensive survey and performance evaluation," *Electronics*, vol. 9, no. 8, pp. 1–12, 12. Aug. 2020. doi: 10.3390/electronics9081295.

[41] F. H. Fadel and S. F. Behadili, "A Comparative Study for Supervised Learning Algorithms to Analyze Sentiment Tweets," *Iraqi Journal of Science*, vol. 63, no. 6, pp. 2712–2724, 30. Jun. 2022, doi: 10.24996/ijs.2022.63.6.36.

[42] J. Y. Khan, Md. T. I. Khondaker, S. Afroz, G. Uddin, and A. Iqbal, "A benchmark study of machine learning models for online fake news detection," *Machine Learning with Applications*, vol. 4, p. 100032, Jun. 2021, doi: 10.1016/j.mlwa.2021.100032.

[43] T. Thaher, M. Saheb, H. Turabieh, and H. Chantar, "Intelligent Detection of False Information in Arabic Tweets Utilizing Hybrid Harris Hawks Based Feature Selection and Machine Learning Models," *Symmetry*, vol. 13, no. 4, p. 556, Apr. 2021, doi: 10.3390/sym13040556.

[44] S. F. Sabbeh and S. Y. Baatwah, "Arabic news credibility on Twitter: An enhanced model using hybrid features," *Journal of Theoretical & Applied Information Technology*, vol. 96, no. 8, 2018.

[45] A. I. Kadhim, "Term Weighting for Feature Extraction on Twitter: A Comparison Between BM25 and TF-IDF," *in 2019 International Conference on Advanced Science and Engineering (ICOASE), Zakho - Duhok, Iraq*, pp. 124-128, 30. May. 2019, doi: 10.1109/ICOASE.2019.8723825.

[46] J. K. Rout, K. K. R. Choo, A. K. Dash, S. Bakshi, S. K. Jena, and K. L. Williams, "A model for sentiment and emotion analysis of unstructured social media text," *Electronic Commerce Research*, vol. 18, no. 1, pp. 181–199, Mar. 2018, doi: 10.1007/s10660-017-9257-8.

[47] Aljarah, M. Habib, N. Hijazi, H. Faris, R. Qaddoura, B. Hammo, M. Abushariah, and M. Alfawareh, "Intelligent detection of hate speech in Arabic social network: A machine learning approach," *Journal of Information Science*, vol. 47, no. 4, pp. 483-501, Aug. 2021. https://doi.org/10.1177/016555152091765.

[48] A. Abbasi and A. Derakhti, "An exploratory study on disinformation and fake news associated with the US 2020 presidential election," in *Bright Internet Global Summit 2020*, 2020.