A RESEARCH ON DETECTION OF FAKE NEWS ON SOCIAL MEDIA USING MACHINE LEARNING ALGORITHMS

Preprint · December 2022

3 authors, including:

Doris Chioma
Bournemouth University

Goodluck John
Bournemouth University

A RESEARCH ON DETECTION OF FAKE NEWS ON SOCIAL MEDIA USING MACHINE LEARNING ALGORITHMS

DORIS CHIOMA AGUH
Computing And Informatics
Bournemouth University
Poole, United Kingdom
s5334566@bournemouth.ac.uk

GOODLUCK OBIANKE JOHN
Computing And Informatics
Bournemouth University
Poole, United Kingdom
s5327090@bournemouth.ac.uk

OYENIKE SARAH AKANJI
Computing And Informatics
Bournemouth University
Poole, United Kingdom
s5527734@bournemouth.ac.uk

ABSTRACT

Both benefits and drawbacks can be associated with using social media for news consumption. On the one hand, social media is used by people as a news source because it is cost-free, easily reachable, and spreads information quickly. On the other hand, it promotes the widespread dissemination of "fake news," or news that intentionally contains incorrect information. The widespread dissemination of fake news could have a negative effect on society and its people. Due to this, identifying fake news on social media has become a very hot research topic recently. Social media's distinct features and barriers to fake news detection make traditional news media detection algorithms ineffective or irrelevant. Fake news is difficult and nontrivial to identify based solely on news content because it is intentionally designed to persuade readers to believe false information. We must therefore consider auxiliary information to support our analysis, such as user social engagements on social media. The large, inaccurate, chaotic, and noisy data generated by consumers' social interactions with fake news makes it challenging to use this additional data. In this study, we in-depth investigate the problem of detecting fake news on social media, including definitions of fake news based on psychological and social theories, modern data mining techniques, evaluation standards, and sample datasets. We also talked about related research areas and potential lines of inquiry for future work on the identification of fake news on social media. Additionally, using the dataset from the Kaggle repository, we deployed four machine learning algorithms: gradient boosting, decision trees, random forests, and logistic regression. From our analysis, Gradient boosting proved to be the best model, with an accuracy score of 100%.

Keywords: Machine Learning, Fake News, Logistic Regression, Random Forest, Gradient Booting, Decision Tree, Classifiers, social media, Detection.

1. INTRODUCTION

More people are likely to seek out and consume news from social media rather than more traditional news sources as more of our lives are spent online interacting with others via social media platforms [10]. The reasons for this shift in consumption patterns are inherent to those social media platforms: reading news on social media is frequently more immediate and less expensive than doing it through traditional journalism, like newspapers or television. Additionally, it may be less complicated to share, talk about, and continue a conversation about the news with friends or other social media users.

5.9% of the world's population, or 4.70 billion people, use social media, with 227 million new users joining in the previous year, according to a Data reportal [11]. In recent years, social media has surpassed television as the main news source. The quality of stories on social media is lower than that of traditional news organizations, despite the benefits social media provides. Fake news, however, is created online for a variety of reasons, such as for commercial and political gain. News is easily and quickly promoted on social media, and it is inexpensive to distribute news online. For instance, during elections, fake news production typically increases in most nations, which frequently results in cases of people dying and so much violence as a result of online rumors.

Fake news, as defined by Cambridge Dictionary [12], is untrue information that is frequently sensational and is produced to be widely shared or distributed in order to generate income or to either promote or denigrate a public figure, political movement, business, etc. In general, spreading false information has the potential to harm both individuals and society. First of all, inaccurate information can affect how people perceive and react to crucial information. Fake news, for instance, can be produced solely to sow doubt and ambiguity in people's minds, making it challenging for them to tell the difference between what is true and false.

Secondly, fake news has the potential to undermine the credibility of the news media. For instance, it was evident that during the 2016 presidential elections in the U.S, social media networks, rather than traditional news outlets, were the primary source of the most widely disseminated false information. Thirdly, false information may persuade readers to readily believe untrue stories. Propagandists frequently use fake news to disseminate false information and influence political outcomes. False news is frequently published with numerous grammatical errors, usually using attention-getting words, a news format, and additional click-bait. Although they seem too good to be true, their sources are frequently not genuine [1].

Tools that can automatically identify false information when it is posted on any social media network must be created in order to reduce the harm misinformation causes on all spheres. Some websites today disseminate untrue information that makes claims to be true while being false or partially true. Social media is frequently used by them as their main tool to generate unauthorized traffic to their websites or social media platforms for their own personal gain. Since this is a global problem, it must be addressed on a global scale. However, since the dawn of time, people have employed a wide range of tools to expedite a variety of tasks.

Due to human innovations, various machines were created, allowing people to satisfy various needs for their daily lives, including travel, industry, and computing. Machine learning is one of these technologies that made has life easier for people [4].

1.1 PROBLEM STATEMENT

Users find online platforms convenient because they can access their messages with ease. The issue is that this enables online criminals to use these platforms to spread false information. This message might be detrimental to both people and society. After reading the news, readers begin to believe it without checking. Fake news is difficult to identify because it is not simple [10].

If fake news is not detected on time, it may spread, and everyone may begin to believe it. Fake news can have an impact on people, groups, or political parties. For instance, false information about the 2016 U.S. presidential election affected people's perceptions and choices [65].

1.2 RESEARCH AIMS AND OBJECTIVES

This work aims to create a system or model that can accurately predict fake news using data from previous news reports.

OBJECTIVES

- To use Python to develop a machine learning model that can accurately detect if some news is fake or real.
- To implement and evaluate which machine learning algorithm is accurate for the detection of fake news
- To evaluate the proposed model against standard evaluation metrics.

1.3 RESEARCH QUESTIONS

1. What is machine learning?
2. Why is machine learning necessary to identify fake news?
3. Which machine learning algorithms are effective in identifying fake news?
4. How are algorithms for machine learning trained to detect false news

2. LITERATURE REVIEW

2.1 FAKE NEWS

Fake news has existed for a very long time—almost as long as it took for information to start being widely disseminated after the printing press was invented in 1943. However, there is no conclusive definition of "fake news," [20].

According to [14], the term "fake" is used to describe articles of news that are intentionally and demonstrably untrue and may lead readers astray. This concept has two essential components: authenticity and aim. First, false information that can be proven wrong is included in the false news. Second, incorrect information is pinpointed with the goal of misleading readers. In recent studies, this definition has been widely used [15,16].

When false information superficially resembles news media content but has a different organizational structure or objective, it is sometimes referred to as "fake news." Additionally, fake news sources frequently lack the editorial guidelines and procedures used by news organizations to verify the accuracy and veracity of information [17].

2.1.1 FAKE NEWS ON TRADITIONAL NEWS MEDIA

Fake news is a long-standing problem. Fake news media ecology has changed over time, moving from newsprint to radio and television to, more recently, online news and social media. "Classical fake news" is the term we use to describe the fake news issue before social media had a significant impact on its production and dissemination. Using psychological and social information ecosystem levels, the effects of fake news on people and the social information ecosystem will be discussed [10].

A. Fake News Psychological Effect:

Real and fake news is difficult for humans to distinguish by nature. This phenomenon and the influence of fake news on public opinion can be explained by a number of psychological and cognitive theories. Consumers are primarily targeted by traditional fake news because of individual weaknesses, and consumers are naturally more susceptible to fake news due to two main factors:

I. **Naive Realism:** They tend to believe that their perceptions of reality are the only valid views and that anyone who disagrees is uninformed, irrational, or biased [18]. II.

II. **Confirmation Bias:** They favor information that confirms the opinions they already have [13].

Consumers frequently mistake fake news for the real thing because of certain cognitive flaws that are inherent. Furthermore, it is very challenging to change a misperception once it has been formed. According to psychological research, correcting misleading information (such as fake news) with the presentation of accurate, factual information does nothing to assist or lessen erroneous views. Still, it sometimes makes them worse, especially among people with cultural beliefs in particular. [21].

B. Social Foundations of the Fake News Ecosystem

When we take into account the system of news consumption, we may also characterize some of the social processes that contribute to the dissemination of false information. According to prospect theory, people make decisions based on the relative gains and losses compared to their current circumstances [19; 22]. This desire to maximize the rewards of a choice also applies to social benefits, such as continued acceptance by people in one's immediate social circle. This desire to maximize a choice's advantages also relates to social advantages, like upholding one's reputation within one's present social network. According to social identity theory [23; 24] and normative influence theory [25; 27], users are likely to choose "socially safe" options when consuming and spreading news information, sticking to the norms established in the community even if the news being shared is fake news. It is crucial to a person's identity and self-esteem that they have this preference for social approval and affirmation.

The cycle of a two-player strategy game for news production and consumption can be used to model this logical theory of interactions with fake news from the perspective of economic game theory [26]. We assume that the information ecosystem has two different types of key players: publishers and consumers, in order to better understand fake news. News reporting is modelled as a mapping from the original signal(s) to the news report(a), with bias distortion(b) as an effect. In other words, b = [1, 0, 1] denotes [left, no, right] biases have an impact on the news publishing process.

This explains how a news article may be biased or distorted to present false information. From two perspectives, the publisher can benefit: I: Short term benefits: This is the concession to maximize profit, which is positively correlated with the number of consumers reached, is of immediate importance. Long-term benefits: this refers to their standing as reliable sources of news. Consumer is divided into two categories: (I): Information benefits: this refers to consumers obtaining accurate and objective information (typically requiring additional investment cost); (II): Intellectual benefits: this refers to receiving news that corresponds with their preexisting beliefs and social expectations, such as confirmation bias and prospect theory. Fake news occurs when a publisher's short-term benefits outweigh its overall benefits, and a consumer's Intellectual benefits outweighs its general benefits while maintaining equilibrium. This discusses the social factors contributing to an information ecology where fake news can thrive.

2.1.2 FAKE NEWS ON SOCIAL MEDIA

This simply refers to some distinguishing characteristics of fake news on social media, emphasizing the salient features of fake news made possible by social media. Keep in mind that social media can be used to disseminate false information as well because it exhibits many of the same characteristics as conventional fake news [10].

A. MALICIOUS SOCIAL MEDIA ACCOUNTS THAT SPREAD PROPAGANDA

If social media has a large number of authorized users, it's possible that some of them are venomous or even fake people. Due to the low cost of opening a social media account, trolls, social bots, and other venomous users are also encouraged to open accounts. A "social bot," is a computer algorithm that runs social media accounts, automatically generates content, and interacts with users on the platform (both humans and other bots) [28]. The spreading of false information on social media and social media manipulation are just two instances where social bots have turned hostile and spiteful.

The 2016 U.S. presidential election was significantly impacted by social bots, according to studies [29], and in the week before the election, roughly 19 million bot accounts tweeted in favor of either Trump or Clinton. The spread of false information on social media is also significantly aided by trolls, actual people who attempt to disturb online communities and elicit an emotional response from users. For instance, data suggests that 1,000 hired Russian trolls disseminated false information about Hillary Clinton.

The context of online discussions and people's moods have a significant impact on Trolling behaviors, making it simple to spread false information among otherwise "normal" online communities [30]. Trolling awakens suppressed negative emotions like anger and fear in people, which breeds distrust and unruly behavior. Lastly, users who are clones can disseminate fake news in a way that combines automated processes with inputs from humans. Typically, people set up automated programs to perform tasks on social media and create clone accounts to act as a disguise. Due to the seamless transition between human and bot functionalities, clone users have a special opportunity to spread false information [31]. Simply put, these harmful social media accounts that are extremely active and biased serve as powerful sources and channels of dissemination.

B. RESONATING CHAMBER EFFECT

Social media provides a new pattern of information obtained and consumed for users. The process of obtaining and consuming information is transitioning from being mediated (by Journalists, for example) to being less mediated [32]. Due to how news feeds appear on users' social media homepages, consumers are only exposed to types of content, which exacerbates the psychological barriers to identifying and avoiding fake news. Users of Facebook, for instance, always follow individuals who share their views, which allows them to receive information that supports their preferred existing narratives [33].

As a result, on social media, users frequently create groups of like-minded individuals where they polarize their views, creating an echo chamber effect. The following psychological elements contribute to the resonating chamber effect by making it simpler for people to take in and accept false information [36]: (I) social Integrity refers In particular a situation when there is not enough information to determine whether the source is telling the truth, people are more likely to believe a source when others believe it to be accurate; and (ii) objective pitch, which states that people may favor information they frequently hear.

Studies have shown that increased exposure to an idea is enough to generate a favorable opinion [35; 37], and users continue to share and consume the same information in echo chambers. As a result, this echo chamber effect creates segmented, homogeneous groups with a relatively limited information ecosystem. Research reveals that homogenous communities become the dominant driver of information diffusion, further promoting contradictions [34].

2.1.3 PROBLEM AREAS RELATING TO FAKE NEWS

A. RUMOUR CLASSIFICATION

A rumour is typically defined as "a piece of spreading information whose authenticity has not yet been established at the time of spreading" [54]. The purpose of a rumour is to provide a context in an ambiguous situation, and its integrity can be true, false, or unverified. Rumour tracking, stance classification, rumour detection, and veracity classification are the four subtasks that are the focus of prior rumour analysis methodologies [54].

To be more specific, rumor detection aims to classify information as rumor or non-rumor [55; 56], rumor tracking aims to gather and filter posts discussing specific stories, rumor stance classification establishes how each relevant post is oriented concerning the rumor's integrity, and authenticity classification aims to predict the rumor's actual truth value. The task that is most closely related to the detection of fake news is the classification of rumor authenticity. Classifying the authenticity of rumors heavily relies on the other functions, necessitating the extraction of positions or viewpoints from relevant posts. The rumor's validity can be ascertained using these posts, which are regarded as significant signals. Fake news is information that is explicitly relevant to current events in the news that may be proven to be untrue. This differs from rumors, which can refer to long-term rumors like conspiracy theories and short-term emerging rumors.

B. TRUTH DISCOVERY

Truth discovery is the process of identifying the real story among conflicting information from various sources [57]. Instead of directly examining the truthful information, truth discovery techniques rely on a collection of contradictory sources that list the properties of objects to determine the truth value. Truth-finding aims to simultaneously evaluate the object's integrity and the source's dependability. The false news detection problem can benefit from a variety of truth-discovery techniques in a variety of situations.

First, modeling the dependability of various news sources can be used to work out the integrity of reported news. Relevant social media posts can also be modelled as social response sources to evaluate the veracity of claims [58; 59]. There are particular additional considerations that must be considered in order to apply truth discovery to the identification of fake news in social media contexts. The majority of truth discovery techniques now focus on managing structured input in the form of Subject-Predicate-Object (SPO) tuples, which is in contrast to social media data, and is primarily unstructured and noisy. Second, truth-discovery techniques are ineffective when only a few news outlets publish a piece of fake news because there aren't enough posts on social media that are pertinent to the piece to serve as additional sources.

C. MISLEADING HEADLINES OR CLICKBAIT DETECTION

In the context of online media, captivating and baiter headlines are frequently referred to as "misleading headlines or clickbait." When reading misleading headlines to arouse their interest, readers are more likely to click the target link. The language used in baiter messages, linked websites, and tweet metadata is used by current misleading headlines or clickbait detection techniques [61, 62, 63]. There are various types of misleading or clickbait headlines, some of which are closely related to false statements [60]. Increasing click-through rates and the subsequent revenue from advertising is frequently clickbait's main objective. The body of clickbait articles is frequently haphazardly structured and inadequately supported as a result. In order to identify fake news articles, researchers have taken advantage of this discrepancy to spot the inconsistencies between headlines and news content [21]. While not all false news may have clickbait headlines, some clickbait headlines may be a key indicator, and there are a number of characteristics that can be used to spot false news.

D. SPAMMER AND BOT DETECTION

Recently, there has been a lot of interest in spammer detection on social media, which seeks to identify bad actors who work together to commit various assaults, including distributing malware, promoting advertisements, spreading pornography, and engaging in stealing of users data, personal login details and credit card numbers[44]. Most methods currently used to identify social spammers are based on variables extracted from user activity and social network data [35; 95; 33; 34]. Because social bots automatically retweet items without vetting the content for accuracy, their prevalence has accelerated the spread of false information [23].

The major challenge with social bots is that they can falsely suggest content and that content is widely supported and liked, which encourages the resonating chamber effect and the spread of false information. Crowdsourcing, social network data, and discriminative features have all been used as foundations for earlier bot detection techniques [23, 55, 54]. As a result, spammers and social media bots may both provide information about specific problematic social media accounts that can be used to spot fake news.

2.2 SOCIAL MEDIA

Social media is a term for computer-based technology that facilitates communication among individuals via online networks and communities. Social media platforms based on the internet enable quick electronic communication of content, including documents, videos, photos, and personal information. On a computer, tablet, or smartphone, users interact with social media through web-based software or applications. [67]

The original purpose of social media was to connect people with friends and family, but businesses later adopted it as a way to reach out to customers through a popular new channel of communication. The advantage of social media is its capacity to connect and converse with anyone on Earth or with a large group of people at once. Social media is used by more than 3.8 billion people worldwide. Social media is a field that is constantly changing and evolving, with new apps like TikTok and Clubhouse joining the ranks of well-established social networks like Facebook, YouTube, Twitter, and Instagram seemingly every year. By 2023, there should be about 257 million social media users in the US, according to predictions [67]. According to the Pew Research Center, social media users are usually younger. Nearly 90% of people between the ages of 18 and 29 used social media in some way. Additionally, these users frequently hold advanced degrees and make over $75,000 a year, making them relatively prosperous [68].

TYPES OF SOCIAL MEDIA

The way that people interact or communicate online has changed as a result of social media. It allows people to communicate with one another, stay in touch with friends who live far away, keep up with current events on a global scale, and access a vast amount of knowledge. People can now connect with one another online and make the world seem more approachable thanks to social media in many different ways. According to a study by [68], using social media is linked to having more friends and a more varied personal network, especially in emerging economies. A lot of teen friendships can start virtually; 57% of teenagers say they have made an online friend.

Even though social media has benefits, many criticize it and draw attention to its negative aspects, likening excessive use to an addiction. It can cause distraction, stress, and jealousy in some people. According to the National Center for Biotechnology Information, social media use is associated with depression. Furthermore, misinformation and false information could circulate on social media. It has been extensively documented how the platform's capacity to spread false information during the 2016 American presidential election affected the outcome. A situation like this utilizes social media and makes it possible for anyone to reach millions of people with erroneous or incomplete content [67].

Despite the fact that there are many different kinds of social media, a method of identifying them is based on the two essential components of social media: media research and social processes. Social networking sites are one of the most well-liked forms of social media; these are programs that let users connect by building personal information profiles, granting friends and coworkers access to those profiles, and sending emails and instant messages to one another. For instance, to name a few, Facebook, Twitter, Instagram, TikTok, and Snapchat. These individual profiles may contain any kind of data, including blogs, images, videos, and audio files. A number of businesses already make use of social networking sites to aid in the development of brand communities [44] or for marketing research in the context of netnography [43,45].

2.3 WHAT IS MACHINE LEARNING?

Artificial intelligence (AI) systems are used to carry out complicated tasks in a similar fashion to how people solve issues. An area of Artificial Intelligence known as Machine Learning is the ability for a machine to mimic intelligent human behaviour. Machine learning has become the preferred method for developing different valuable software system for applications, for example, robot control, audio recognition and computer vision amongst others. AI system engineers are now aware that, for many tasks, it may be more difficult to manually programme a system by assuming the desired response for every possible input than to train a system by providing instances of proper input and output behaviour.

According to [4], a machine learning system's function can be either descriptive or predictive. This implies it can either utilise the data to explain what happened or forecast what will happen or the system uses the information to advise a course of action.

2.3.1 CATEGORIES OF MACHINE LEARNING

There are four categories of machine learning.

A. SUPERVISED MACHINE LEARNING:

This form of Machine learning requires the use of labelled datasets to train models, which enables the models to improve their accuracy over time. For example, the computer could learn how to recognize images of dogs on its own after being trained with images of dogs and other items that have all been recognized by humans. Supervised machine learning is now considered to be the most popular type of machine learning [4].

B. UNSUPERVISED MACHINE LEARNING:

In contrast to Supervised Machine learning, this requires examining unlabeled data for patterns. Unsupervised machine learning can spot patterns or trends that people are not actively looking for. For instance, it may identify the various customer groups by looking at online sales data. [4].

C. REINFORCEMENT LEARNING: This is the type of machine learning that closely relates to how people learn. The deployed agent picks up new information by interacting with its environment and experiencing positive or negative consequences. Examples of these algorithms are the temporal difference, deep adversarial networks, and Q-learning [6].

D. SEMI-SUPERVISED MACHINE LEARNING: This is a combination of supervised and unsupervised machine learning techniques. It can be useful in data mining applications with existing unlabeled data, that requires a considerable amount of time to transform to the labelled data. With more prevalent supervised machine learning techniques, you train an algorithm using a dataset that has been "labelled" and contains outcome data for each record. [13].

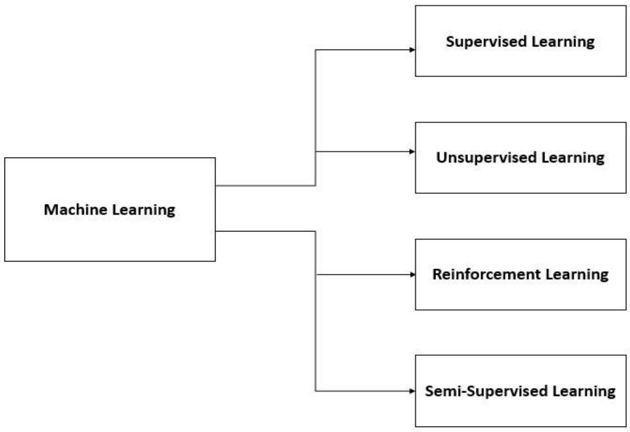

Figure 1: Categories of Machine Learning

2.4 RELATED STUDIES

Several research using different approaches to fake news detection has taken place over the years. According to [8] in his study, used machine learning, natural language processing, and artificial intelligence ideas to conduct binary classification on a variety of online news items to enable the user the ability to judge if a news report is accurate or not and to verify the legitimacy of the publishing website. Based on user input, the model determines whether the news is authentic or fake. The model was trained on a dataset using various NLP and machine learning approaches, and several performance indicators were employed to evaluate its effectiveness. With a 65% accuracy rate, the results indicated that Logistic Regression had the highest accuracy. Grid search parameter optimization was used to enhance the performance of logistic Regression further, and the accuracy reached 75%.

On the other hand, POLITIFACT.COM's LIAR dataset was used by [9]. The LIAR dataset was used to predict false news more accurately, utilizing model ensemble techniques. An effort was also made to simplify the issue statement to binary classification and apply the same ensemble methodologies to provide a more practical and effective method for exact computation. The results showed a combination of Bagging Classifier and AdaBoost after 150 iterations achieved 70% accuracy in precision, F1-Score, and recall.

Using the same dataset (LIAR) in another study, [7] analyzed the reflections on detecting fake news and looked at traditional machine learning models. To test and select the best features to obtain the highest precision, according to the results of the confusion matrix, research of feature selection methods was done on the LIAR dataset. The six detection algorithms employed are XGboost, Random Forests, Naive Bayes, K-Nearest Neighbors (KNN), Decision Tree, and SVM. The result revealed that XGboost had the best accuracy, at over 75%, followed by SVM and Random Forest, both of which had accuracy levels of approximately 73%.

There are several textual characteristics that can be utilized to differentiate between real and false information, as was examined by [2].The performance of four real-world datasets was evaluated by training various machine learning algorithms using those properties and different ensemble methods. The result of the experiment demonstrates that our ensemble learner technique outperforms individual learners.

A hybrid approach that combines the linguistic features of language and the network analysis approach was presented by [38]. This method is not always suitable as network information may be limited or unavailable. Mihalcea and Stratparava [40], Showed that deep learning makes it possible, to some extent, to distinguish between false and true information. Feng et al. [41] applied syntactic stylometry to texts, allowing misleading texts to be classified by finding statistical or syntactic patterns. Text analysis is an essential resource for detecting fake news because text analysis techniques are well-known. A linguistic taxonomy was developed by [39] for detecting fake news, suggesting that linguistic features are a more critical factor in seeing fake news than real news. These approaches lean heavily toward language-based analysis and have some limitations. To overcome this, we need to combine other features related to messages. Combining and integrating Google metadata improved her classification by 3% with her F1 score for the six-label classification problem

3. RESEARCH METHODOLOGY

The concept of Big Data has made it possible to have a massive amount of data, including news. However, this often comes in an unstructured format, indicating that it lacks an immediately discernible structure and does not follow a data model, making it difficult for computer programs to use. Making computers capable of processing or interpreting human language is the goal of the Artificial Intelligence (AI) discipline known as Natural Language Processing (NLP). Theoretically, NLP achieves this by instructing computers to examine and assess enormous amounts of natural language data. With processing power made possible by diverse software, it can transform speech and text. The building blocks are data set and machine learning algorithms. Computers can be taught to distinguish between real and fake news accurately using NLP. Machine learning algorithms such as Logistic Regression classifier, Decision Tree Classifier, Random Forest Classifier and Gradient Boosting Classifier are used for classification in this model.

The model is constructed in two segments. Firstly, the model is evaluated and trained using four distinct classifiers; after that, the classifiers are compared to determine which performed best and is suitable for predicting fake news. The second section takes the user's keyword or text and determines whether the news's integrity is true or fake. Python and its Sci-kit libraries have been utilized, and Python has a wide range of libraries and add-ons that make it simple to use in machine learning. The Scikit-learn library is an excellent resource to get various machine learning algorithms, making them widely accessible for Python and offers a quick and straightforward assessment of Machine learning techniques [8].

3.1 WORKFLOW

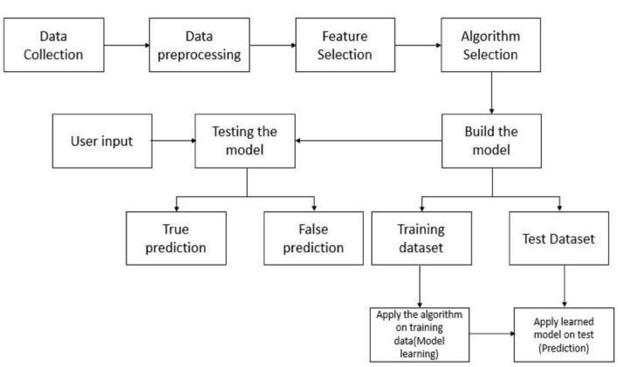

Figure 2: Workflow

A. DATA COLLECTION:

Two datasets were downloaded from Kaggle repository ‘Fake.csv.’ and ‘True.csv’[66]. Each dataset consists of four columns: Title, Text, Subject, and Date. The first, second, third and fourth column describes the headline, body of the article, the type of article, and date of publication respectively. The dataset is 23481 * 4 (23481 rows and four columns) and 21417 * 4 (21417 rows and four columns) for ‘Fake’ and ‘True’, respectively.

B. PREPROCESSING THE TEXT

Social media data usually comes in an unstructured format, as communication is mainly in natural human languages. Hence, we must transform our dataset into a design recognizable by the computer for processing in our predictive model. The below steps were deployed

- Firstly, we inserted a column called ‘Class’ as our target feature in the dataset, 0 for ‘Fake’ and 1 for ‘True.’
- The last ten rows from both files were merged to form the test dataset, namely ‘manual testing
- The two datasets,’ Fake’ and ‘True’, were combined to form the training dataset, 44878 * 4
- Irrelevant columns were removed, leaving us with the most critical columns for our training, ‘Text’ and ‘Class.’

C. DATA CLEANING

It is important to highlight the characteristics of the text data the machine learning will detect and process, hence data cleaning is done. Here, we defined a function to remove punctuation and memorable characters as the features created from a body of words have a significant impact on how well a text classification model performs. Common terms like "a," "the," and "them," among others, are eliminated throughout this process, leaving only words that appear at least a certain number of times in each text dataset.

D. FEATURE EXTRACTION

It is important to highlight the characteristics of the text data the machine learning will detect and process, hence data cleaning is done. Here, we defined a function to remove punctuation and memorable characters as the features created from a body of words have a significant impact on how well a text classification model performs. Common terms like "a," "the," and "them," among others, are eliminated throughout this process, leaving only words that appear at least a certain number of times in each text dataset.

Another way is Hashing Vectorizer, also a method provided by Scikit-learn. It creates a matrix of token occurrences from a collection of text documents, just like Count Vectorizer, but the process is slightly different. TF-IDF vectorizer, another great tool provided by the scikit-learn library and a standard method for conventional machine learning methods [46], was used in this model. It is a method that converts text into a significant representation of numbers, which is suitable for machine prediction algorithms. Tokenizing documents, learning vocabulary and inverse document frequency weightings, and encoding documents are all features of the TfIdfVectorizer. You could also just calculate the inverse document frequencies and begin encoding documents if you already have a learnt Count Vectorizer that you can use with a Tfidf Transformer. The purpose of TF-IDF is to highlight frequent words in documents but not across documents.
TF-IDF is a combination of two terms:[47]

- TF (Term Frequency) — It describes the frequency of a word in a sentence.
- IDF (Inverse Document Frequency) is known as the naturally occurring log of a number of comprehensive texts divided by the documents where the word appears.

3.2 CLASSIFICATION

The training of the classifiers is discussed in this section. The class of the text (True or False) was predicted using various classifiers. We specifically explored Decision Tree Classification, Logistic Regression, Random Forest Classifier and Gradient Boosting Classifier. These classifiers were implemented using Scikit-learn library, a Python module.

A. DECISION TREE CLASSIFIER

A Decision Tree (D-Tree) is an algorithm in machine learning under the supervised category that makes judgments using rules, much like humans do. A Decision Tree is used to construct a training model that may be used to predict the class or value of the target variable by learning simple decision rules obtained from training data. Each node in a D-Tree represents a test over an attribute value, each branch represents the result of the test operation, and the tree's leaf nodes represent classes. This structure is like a flowchart. The selection process can quickly construct intelligible rules and perform classification [49]. The decision tree is vulnerable to errors in multi-class situations and with smaller training cases. It is crucial to keep in mind that no root-to-leaf path should possess the same discrete property twice. [50, 51]. This is made using a parameter in the Scikit-learn library’s Python implementation called “criterion”. Using this option, “Gini” or “Entropy” function can be selected to evaluate the quality of the split. It is however important to select the right criteria because a tree’s accuracy is significantly impacted by the choice of strategic splits. For example, Regression and Classification trees have different decision criteria.

Decision Tree Pseudo-code

```
Generate Decision Tree (Sample s, features F)

  1. If stop _conditions (S, F) = true then

     leaf = create_Node () Leaf.lable= classify(s)

     Return leaf

  2. root = create_Node ()
  3. root. Testcondition = find_bestSplit(s,f)
  4. v = {v }l v a possible outcome of root.testcondition)
  5. for each value v∈ V:
  6. sv: = {s | root. Testcondition(s) = v and s∈ S};
  7. child = Tree_Growth (Sv, F) ;
  8. Grow child as a descent of roof and label the edge (root→child) as v
     Return root
```

B. LOGISTIC REGRESSION

A statistical technique called Logistic Regression is used to build machine learning models with a dichotomous dependent variable. With nominal, ordinal, or interval data types as independent variables, Logistic Regression can be used to describe the relationship between one dependent variable and one or more independent variables. It is worthy to note that method analyses and predicts the chance of an incident occurring by comparing data to a logit function because it is a classification algorithm rather than a regression technique. Logistic Regression offers some advantages making it one of the preferred classifiers for machine learning predictions. Firstly, implementing and training a model using logistic regression is simple; Also, due to its excellent interpretability, it does not require a lot of computational resources. Furthermore, it evaluates a coefficient's significance and the direction of any association, whether it be positive or negative [64].

Pseudo code for Logistic Regression

Input: Training data

1. For i← 1 to k
2. For each training data instance d_i:
3. Set the target value for the regression to
   $z_i \leftarrow \dfrac{y_j - P(1 \mid d_j)}{[P(1 \mid d_j) \cdot (1 - P(1 \mid d_j))]}$
4. nitialize the weight of instance d_j to $P(1 \mid d_j) \cdot (1 - P(1 \mid d_j))$
5. finalize a f(j) to the data with class value (z_j) & weights (wj)
   Classification Label Decision
6. Assign (class label:1) if $P(1|d_j) > 0.5$, otherwise (class label: 2)

C. RANDOM FOREST

A Random Forest consists of numerous separate decision trees that work together as an ensemble. From each tree, a class prediction is made and the class with the highest rating is chosen as the model's prediction. After fitting numerous decision tree classifiers to various dataset samples, a Random Forest acts as a meta-estimator that uses averaging to improve projected accuracy and decrease overfitting. The reason the random forest model performs well is because individual constituent models do not perform better than a committee of many largely uncorrelated models (trees) working together. Each tree is built using features and bagging randomness, which results in having an uncorrelated forest of trees whose prediction by committee is more accurate than any one tree.

Random Forest Pseudo-code

To make *n* classifiers:
For *naïve* = 1 to *n* **do**
Sample the training data T randomly with replacement for Ti|output Build a Ti-containing root node, Ni
Call BuildTree (*Ni*)
end For
BuildTree (N):
If N includes instances of only one class, then returns
else

Select z% of the possible splitting characteristics at random in N
Select the feature *F* with the highest information gain to split on
Create f child nodes of *N*, *Ni* ,..., *Nf* , where *F* has *f* possible values (*F*1, ... , *Ff*)

For *naïve* = 1 to *f* **do**
Set the contents of *Ni* to *Ti*, where *Ti* is all instances in *N* that match *Fi* Call BuildTree (*Ni*)

end for end if

D. GRADIENT BOOSTING CLASSIFIER

The Gradient Booting algorithm has produced the best results in machine learning solutions for business [48], particularly with large and complex datasets. It stands out for its prediction speed and accuracy and helps to minimize the bias error of the model. The gradient boost algorithm builds models sequentially; hence, subsequent models attempt to minimize the errors of the preceding model by building a new model on the errors of the earlier model.

Pseudocode for Gradient Boosting classifier

$$\hat{f}(x) = \sum_{i=0}^{t-1} \hat{f}_i(x),$$

$$r_{it} = -\left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}\right]_{f(x)=\hat{f}(x)}, \quad \text{for } i = 1, \ldots, n,$$

$$\theta_t = \arg\min_{\theta} \sum_{i=1}^{n} \left(r_{it} - h(x_i, \theta)\right)^2,$$

$$\rho_t = \arg\min_{\rho} \sum_{i=1}^{n} L(y_i, \hat{f}(x_i) + \rho \cdot h(x_i, \theta_t))$$

3.3 PERFORMANCE EVALUATION METRIC OF THE CLASSIFIERS

It is necessary to develop certain performance metrics that can be utilised to assess the quality of any classifier that is being considered to evaluate any model's performance properly. The effectiveness of classifiers has been evaluated in this study using five different performance indicators [39, 49]. The following measures have been briefly discussed:

A. Confusion Matrix (CM)

The confusion matrix is a table that displays a condensed evaluation of a classification model's performance. The diagonal entries are those for which the learning method produces the desired outcomes.

- **True Positive (TP**) instances are training examples for which both our hypotheses and the true class are positive. They are what are known as real positives.
- **False Positive (FP):** This refers to the learning algorithm incorrectly identifying samples that are truly negative as positive.
- **True Negative (TN**) examples are those training examples for which we hypothesized as negative, and the true class is negative. They qualify as real negatives.
- **False Negative (FN):** The learning system incorrectly classifies the samples as negative when they are truly positive.

By defining this as a classification model, the following metrics can be defined:

B. PRECISION

The exactness of a classifier can be gauged by its precision. For each class, it describes the ratio of true positives to the total of true and false positives. In other words, what proportion of all incidents that were classified as positive were accurate? [52].

$$\text{Precision} = \frac{|TP|}{|TP| + |FP|}$$

C. RECALL

Recall is a metric for how well-rounded a classifier is; it measures how well it can identify all cases that are classified as positive. For each class, it is the ratio of true positives to the sum of true positives and false negatives. What percentage of occurrences that were accurately labelled were all instances that were genuinely positive, to put it another way? [52].

$$\text{Recall} = \frac{|TP|}{|TP| + |FN|}$$

D. F1 SCORE

A weighted harmonic mean of recall and precision makes up the F1 score, with 1.0 representing the best result and 0.0 the lowest. Because F1 scores factor precision and recall into their computation, they often perform worse than accuracy measurements. As a result, rather than focusing on overall accuracy, it is often recommended to evaluate classifier models using the weighted average of F1[52].

$$\text{F1 Score} = \frac{\text{precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

E. ACCURACY

Accuracy (ACC) is calculated as the total number of correct predictions divided by the total size of the dataset. ACC ranges from 0.0 to 1.0, with 1.0 being the best [52].

$$\text{Accuracy} = \frac{|TP| + |TN|}{|TP| + |TN| + |FP| + |FN|}$$

F. SUPPORT

Support describes the number of real instances of the class in the provided dataset. is known as support. Unequal support in the training data may indicate that the classifier's reported scores have fundamental problems and highlight the necessity of different sampling methods. Support is constant regardless of the model and it diagnoses the evaluation process [53].

The effectiveness of a classifier from several views can be evaluated using the above metrics as they are widely accepted in the machine learning community. While accuracy identifies the similarity between fake news predictions and actual false news, precision measures the proportion of all identified fake news that is classified as such; this is critical to fake news detection. However, given how frequently fake news datasets are skewed, a high degree of precision can be easily achieved by generating fewer pessimistic forecasts. As a result, the proportion of identified fake news that are predicted to be fake news is calculated using recall. F1 score is utilized to combine precision and recall, which can offer a general prediction performance for fake news identification. F1 score is a combination of precision and recall highlighting the general prediction performance for fake news detection. The greater the value of Accuracy, Recall, Precision and F1, the better the performance [52].

4. RESULTS AND IMPLEMENTATION

A. First, we imported the libraries

FAKE NEWS DETECTION USING MACHINE LEARNING ALGORITHM

IMPORTING THE DEPENDENCIES

```
In [1]: import pandas as pd
        import numpy as np
        import seaborn as sns
        import matplotlib.pyplot as plt
        from sklearn.model_selection import train_test_split
        from sklearn.metrics import accuracy_score
        from sklearn.metrics import classification_report
        import re
        import string
```

Figure 3: Imported Libraries.

B. We imported the datasets that we had gotten from Kaggle repository and converted it to a Pandas data frame that python interface can read and printed the first 10 roles of the true and fake datasets. This is shown in the figure below.

IMPORTING THE DATASETS

```
In [2]: df_fake = pd.read_csv("Fake.csv")
        df_true = pd.read_csv("True.csv")

In [3]: df_fake.head(10)

Out[3]:
```

	title	text	subject	date
0	Donald Trump Sends Out Embarrassing New Year'...	Donald Trump just couldn t wish all Americans ...	News	December 31, 2017
1	Drunk Bragging Trump Staffer Started Russian ...	House Intelligence Committee Chairman Devin Nu...	News	December 31, 2017
2	Sheriff David Clarke Becomes An Internet Joke...	On Friday, it was revealed that former Milwauk...	News	December 30, 2017
3	Trump Is So Obsessed He Even Has Obama's Name...	On Christmas day, Donald Trump announced that ...	News	December 29, 2017
4	Pope Francis Just Called Out Donald Trump Dur...	Pope Francis used his annual Christmas Day mes...	News	December 25, 2017
5	Racist Alabama Cops Brutalize Black Boy While...	The number of cases of cops brutalizing and ki...	News	December 25, 2017
6	Fresh Off The Golf Course, Trump Lashes Out A...	Donald Trump spent a good portion of his day a...	News	December 23, 2017
7	Trump Said Some INSANELY Racist Stuff Inside ...	In the wake of yet another court decision that...	News	December 23, 2017
8	Former CIA Director Slams Trump Over UN Bully...	Many people have raised the alarm regarding th...	News	December 22, 2017
9	WATCH: Brand-New Pro-Trump Ad Features So Muc...	Just when you might have thought we d get a br...	News	December 21, 2017

Figure 4: First 10 rolls of Imported Datasets

C. We created a column called class as our target feature and printed the shape and frame of our dataset as shown in figure 5 below.

INSERTING A COLUMN CALLED CLASS AS OUR TARGET FEATURE

```
In [5]: df_fake["class"] = 0
        df_true["class"] = 1

In [6]: df_fake.shape, df_true.shape

Out[6]: ((23481, 5), (21417, 5))
```

Figure 5

D. We removed the last 10 rows of our dataset to be used for manual testing of our models and printed the shape of our datasets after removal. We created the class as our target feature in our manual testing dataset and saved it as manual_testing.csv. as shown in figures 6, 6a, 6b and 6c below.

```
In [7]: # Removing last 10 rows for manual testing
        df_fake_manual_testing = df_fake.tail(10)
        for i in range(23480,23470,-1):
            df_fake.drop([i], axis = 0, inplace = True)


        df_true_manual_testing = df_true.tail(10)
        for i in range(21416,21406,-1):
            df_true.drop([i], axis = 0, inplace = True)

In [8]: df_fake.shape, df_true.shape

Out[8]: ((23471, 5), (21407, 5))
```

Figure 6

```
In [9]: df_fake_manual_testing["class"] = 0
        df_true_manual_testing["class"] = 1

/var/folders/fh/ng8vc5qd1j94hz47jmcd6gv80000gn/T/ipykernel_2432/860779283.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning
-a-view-versus-a-copy
  df_fake_manual_testing["class"] = 0
/var/folders/fh/ng8vc5qd1j94hz47jmcd6gv80000gn/T/ipykernel_2432/860779283.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning
-a-view-versus-a-copy
  df_true_manual_testing["class"] = 1
```

Figure 6a

```
In [10]: df_fake_manual_testing.head(10)
Out[10]:
```

	title	text	subject	date	class
23471	Seven Iranians freed in the prisoner swap have...	21st Century Wire says This week, the historic...	Middle-east	January 20, 2016	0
23472	#Hashtag Hell & The Fake Left	By Dady Chery and Gilbert MercierAll writers ...	Middle-east	January 19, 2016	0
23473	Astroturfing: Journalist Reveals Brainwashing ...	Vic Bishop Waking TimesOur reality is carefull...	Middle-east	January 19, 2016	0
23474	The New American Century: An Era of Fraud	Paul Craig RobertsIn the last years of the 20t...	Middle-east	January 19, 2016	0
23475	Hillary Clinton: 'Israel First' (and no peace ...	Robert Fantina CounterpunchAlthough the United...	Middle-east	January 18, 2016	0
23476	McPain: John McCain Furious That Iran Treated ...	21st Century Wire says As 21WIRE reported earl...	Middle-east	January 16, 2016	0
23477	JUSTICE? Yahoo Settles E-mail Privacy Class-ac...	21st Century Wire says It s a familiar theme. ...	Middle-east	January 16, 2016	0
23478	Sunnistan: US and Allied 'Safe Zone' Plan to T...	Patrick Henningsen 21st Century WireRemember ...	Middle-east	January 15, 2016	0
23479	How to Blow $700 Million: Al Jazeera America F...	21st Century Wire says Al Jazeera America will...	Middle-east	January 14, 2016	0
23480	10 U.S. Navy Sailors Held by Iranian Military ...	21st Century Wire says As 21WIRE predicted in ...	Middle-east	January 12, 2016	0

```
In [11]: df_true_manual_testing.head(10)
```

Figure 6b

```
n [11]: df_true_manual_testing.head(10)
ut[11]:
```

	title	text	subject	date	class
21407	Mata Pires, owner of embattled Brazil builder ...	SAO PAULO (Reuters) - Cesar Mata Pires, the ow...	worldnews	August 22, 2017	1
21408	U.S., North Korea clash at U.N. forum over nuc...	GENEVA (Reuters) - North Korea and the United ...	worldnews	August 22, 2017	1
21409	U.S., North Korea clash at U.N. arms forum on ...	GENEVA (Reuters) - North Korea and the United ...	worldnews	August 22, 2017	1
21410	Headless torso could belong to submarine journ...	COPENHAGEN (Reuters) - Danish police said on T...	worldnews	August 22, 2017	1
21411	North Korea shipments to Syria chemical arms a...	UNITED NATIONS (Reuters) - Two North Korean sh...	worldnews	August 21, 2017	1
21412	'Fully committed' NATO backs new U.S. approach...	BRUSSELS (Reuters) - NATO allies on Tuesday we...	worldnews	August 22, 2017	1
21413	LexisNexis withdrew two products from Chinese ...	LONDON (Reuters) - LexisNexis, a provider of l...	worldnews	August 22, 2017	1
21414	Minsk cultural hub becomes haven from authorities	MINSK (Reuters) - In the shadow of disused Sov...	worldnews	August 22, 2017	1
21415	Vatican upbeat on possibility of Pope Francis ...	MOSCOW (Reuters) - Vatican Secretary of State ...	worldnews	August 22, 2017	1
21416	Indonesia to buy $1.14 billion worth of Russia...	JAKARTA (Reuters) - Indonesia will buy 11 Sukh...	worldnews	August 22, 2017	1

```
n [12]: df_manual_testing = pd.concat([df_fake_manual_testing,df_true_manual_testing], axis = 0)
        df_manual_testing.to_csv("manual_testing.csv")
```

Figure 6c

E. We merged the two datasets to be a single dataset so that our models can perform maximally. Then we dropped the columns we would not need, as shown in figures 7 and 7a below.

Merging True and Fake Dataframes

```
: df_merge = pd.concat([df_fake, df_true], axis =0 )
  df_merge.head(15)
```

	title	text	subject	date	class
0	Donald Trump Sends Out Embarrassing New Year'...	Donald Trump just couldn t wish all Americans ...	News	December 31, 2017	0
1	Drunk Bragging Trump Staffer Started Russian ...	House Intelligence Committee Chairman Devin Nu...	News	December 31, 2017	0
2	Sheriff David Clarke Becomes An Internet Joke...	On Friday, it was revealed that former Milwauk...	News	December 30, 2017	0
3	Trump Is So Obsessed He Even Has Obama's Name...	On Christmas day, Donald Trump announced that ...	News	December 29, 2017	0
4	Pope Francis Just Called Out Donald Trump Dur...	Pope Francis used his annual Christmas Day mes...	News	December 25, 2017	0
5	Racist Alabama Cops Brutalize Black Boy While...	The number of cases of cops brutalizing and ki...	News	December 25, 2017	0
6	Fresh Off The Golf Course, Trump Lashes Out A...	Donald Trump spent a good portion of his day a...	News	December 23, 2017	0
7	Trump Said Some INSANELY Racist Stuff Inside ...	In the wake of yet another court decision that...	News	December 23, 2017	0
8	Former CIA Director Slams Trump Over UN Bully...	Many people have raised the alarm regarding th...	News	December 22, 2017	0
9	WATCH: Brand-New Pro-Trump Ad Features So Muc...	Just when you might have thought we d get a br...	News	December 21, 2017	0
10	Papa John's Founder Retires, Figures Out Raci...	A centerpiece of Donald Trump s campaign, and ...	News	December 21, 2017	0
11	WATCH: Paul Ryan Just Told Us He Doesn't Care...	Republicans are working overtime trying to sel...	News	December 21, 2017	0
12	Bad News For Trump — Mitch McConnell Says No ...	Republicans have had seven years to come up wi...	News	December 21, 2017	0
13	WATCH: Lindsey Graham Trashes Media For Portr...	The media has been talking all day about Trump...	News	December 20, 2017	0
14	Heiress To Disney Empire Knows GOP Scammed Us...	Abigail Disney is an heiress with brass ovarie...	News	December 20, 2017	0

Figure 7

```
In [14]: df_merge.columns
Out[14]: Index(['title', 'text', 'subject', 'date', 'class'], dtype='object')
```

Removing columns which are not required

```
In [15]: df = df_merge.drop(["title", "subject","date"], axis = 1)

In [16]: df.head(15)
Out[16]:
```

	text	class
0	Donald Trump just couldn t wish all Americans ...	0
1	House Intelligence Committee Chairman Devin Nu...	0
2	On Friday, it was revealed that former Milwauk...	0
3	On Christmas day, Donald Trump announced that ...	0
4	Pope Francis used his annual Christmas Day mes...	0
5	The number of cases of cops brutalizing and ki...	0
6	Donald Trump spent a good portion of his day a...	0
7	In the wake of yet another court decision that...	0
8	Many people have raised the alarm regarding th...	0
9	Just when you might have thought we d get a br...	0
10	A centerpiece of Donald Trump s campaign, and ...	0
11	Republicans are working overtime trying to sel...	0

Figure 7a

F. We checked the dataset to find out if there were any null values or strings, and We shuffled our dataset as shown in figures 8, 8a and 8b below.

14	Abigail Disney is an heiress with brass ovarie...	0

```
In [17]: df.isnull().sum()

Out[17]: text     0
         class    0
         dtype: int64
```

Figure 8

Random Shuffling the dataframe

```
In [18]: df = df.sample(frac = 1)

In [19]: df.head(15)
Out[19]:
```

	text	class
15082	Indoctrination by Disney pretty much covers ev...	0
6622	There are many important issues in the world: ...	0
12963	ALGIERS (Reuters) - French President Emmanuel ...	1
21822	1). I m a Christian.2). I m NOT gay.3). He fea...	0
16183	(Note: Readers might find some language offen...	1
18165	CAIRO (Reuters) - Sarah Hegazy has been jailed...	1
19706	Here is Danney Williams tragic story. There is...	0
9584	WASHINGTON (Reuters) - U.S. Republican Nationa...	1
21202	Does anyone remember a time in recent history ...	0
2068	Fox & Friends continues to embarrass themselve...	0
21876	But wait wasn t the Muslim Brotherhood s Arab ...	0
4615	WASHINGTON (Reuters) - The U.S. House voted on...	1
17983	ISTANBUL (Reuters) - Turkey could seek a deal ...	1
8526	(Reuters) - U.S. Republican presidential candi...	1
16232	KATHMANDU (Reuters) - A bus carrying passenger...	1

Figure 8a

```
:n [20]: df.reset_index(inplace = True)
         df.drop(["index"], axis = 1, inplace = True)

:n [21]: df.columns

)ut[21]: Index(['text', 'class'], dtype='object')

:n [22]: df.head(10)

)ut[22]:
```

	text	class
0	Indoctrination by Disney pretty much covers ev...	0
1	There are many important issues in the world: ...	0
2	ALGIERS (Reuters) - French President Emmanuel ...	1
3	1). I m a Christian.2). I m NOT gay.3). He fea...	0
4	(Note: Readers might find some language offen...	1
5	CAIRO (Reuters) - Sarah Hegazy has been jailed...	1
6	Here is Danney Williams tragic story. There is...	0
7	WASHINGTON (Reuters) - U.S. Republican Nationa...	1
8	Does anyone remember a time in recent history ...	0
9	Fox & Friends continues to embarrass themselve...	0

Figure 8b

G. We created a function to clean up our texts further because some tasks in the text can challenge our models and affect the accuracy of results. This is shown in figure 9 below.

Creating a function to process the texts

```
In [23]: def wordrop(text):
             text = text.lower()
             text = re.sub('\[.*?\]', '', text)
             text = re.sub("\\W"," ",text)
             text = re.sub('https?://\S+|www\.\S+', '', text)
             text = re.sub('<.*?>+', '', text)
             text = re.sub('[%s]' % re.escape(string.punctuation), '', text)
             text = re.sub('\n', '', text)
             text = re.sub('\w*\d\w*', '', text)
             return text

In [25]: df["text"] = df["text"].apply(wordrop)
```

Figure 9

H. We defined our independent and dependent variables and then split our dataset into test and training datasets, as shown in the figure 10 below.

Defining the dependent and independent variables

```
n [26]: x = df["text"]
        y = df["class"]
```

Splitting the dataset to Training and Testing sets

```
n [27]: x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.33)
```

Figure 10

I. We converted our texts to vectors using TFIDF vectorizer, as shown in figure 11 below.

Convert text to vectors

```
In [28]: from sklearn.feature_extraction.text import TfidfVectorizer

         vectorization = TfidfVectorizer()
         xv_train = vectorization.fit_transform(x_train)
         xv_test = vectorization.transform(x_test)
```

Figure 11

J. We created our classification models for the different algorithms we have chosen and trained them with 77% of our data set, tested our models with 33% and printed their classification report. These are shown in Figure 12, below

CREATING OUR MODELS

Logistic Regression

```
In [30]: from sklearn.linear_model import LogisticRegression

         LR = LogisticRegression()
         LR.fit(xv_train,y_train)

Out[30]: LogisticRegression()

In [31]: pred_lr=LR.predict(xv_test)

In [32]: LR.score(xv_test, y_test)

Out[32]: 0.986428089128967
```

Figure 12: Logistic Regression

```
In [33]: print(classification_report(y_test, pred_lr))

              precision    recall  f1-score   support

           0       0.99      0.99      0.99      7705
           1       0.98      0.99      0.99      7105

    accuracy                           0.99     14810
   macro avg       0.99      0.99      0.99     14810
weighted avg       0.99      0.99      0.99     14810
```

Figure 12a: Logistic Regression Classification Report.

Decision Tree Classification

```
In [34]: from sklearn.tree import DecisionTreeClassifier

         DT = DecisionTreeClassifier()
         DT.fit(xv_train, y_train)

Out[34]: DecisionTreeClassifier()

In [35]: pred_dt = DT.predict(xv_test)

In [36]: DT.score(xv_test, y_test)

Out[36]: 0.9948008102633356
```

Figure 13: Decision Tree Classifier

```
In [37]: print(classification_report(y_test, pred_dt))

              precision    recall  f1-score   support

           0       0.99      1.00      1.00      7705
           1       1.00      0.99      0.99      7105

    accuracy                           0.99     14810
   macro avg       0.99      0.99      0.99     14810
weighted avg       0.99      0.99      0.99     14810
```

Figure 13a: Decision Tree classification Report.

Random Forest Classifier

```
In [38]: from sklearn.ensemble import RandomForestClassifier

         RFC = RandomForestClassifier(random_state=0)
         RFC.fit(xv_train, y_train)

Out[38]: RandomForestClassifier(random_state=0)

In [39]: pred_rfc = RFC.predict(xv_test)

In [40]: RFC.score(xv_test, y_test)

Out[40]: 0.9911546252532073
```

Figure 14: Random Forest Classifier

```
In [41]: print(classification_report(y_test, pred_rfc))

              precision    recall  f1-score   support

           0       0.99      0.99      0.99      7705
           1       0.99      0.99      0.99      7105

    accuracy                           0.99     14810
   macro avg       0.99      0.99      0.99     14810
weighted avg       0.99      0.99      0.99     14810
```

Figure 14a: Random Forest classification Report

Gradient Boosting Classifier

```
In [42]: from sklearn.ensemble import GradientBoostingClassifier

         GBC = GradientBoostingClassifier(random_state=0)
         GBC.fit(xv_train, y_train)
Out[42]: GradientBoostingClassifier(random_state=0)

In [43]: pred_gbc = GBC.predict(xv_test)

In [44]: GBC.score(xv_test, y_test)

Out[44]: 0.9957461174881836
```

Figure 15: Gradient Boosting Classifier.

```
In [45]: print(classification_report(y_test, pred_gbc))
              precision    recall  f1-score   support

           0       1.00      0.99      1.00      7705
           1       0.99      1.00      1.00      7105

    accuracy                           1.00     14810
   macro avg       1.00      1.00      1.00     14810
weighted avg       1.00      1.00      1.00     14810
```

Figure 15a: Gradient Boosting classification Report

K. We printed the confusion matrix of each classifier to check the accuracy of prediction. This is shown in the figures below.

- Confusion matrix of Logistic Regression Classifier

```
In [101]: from sklearn.metrics import confusion_matrix, ConfusionMatrixDisplay
          from sklearn import metrics

          confusion_matrix = metrics.confusion_matrix(y_test, pred_lr)
          cm_display = metrics.ConfusionMatrixDisplay(confusion_matrix = confusion_matrix,
                                                      display_labels = [True, False])
          cm_display.plot()
          plt.show()
```

Figure 16

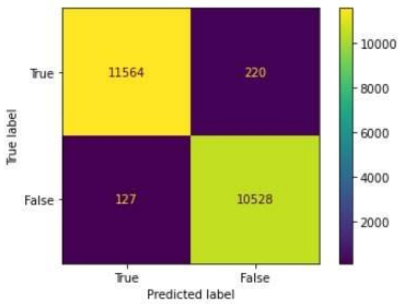

Figure 16a

- Confusion matrix of Decision Tree Classifier

```
In [102]: confusion_matrix = metrics.confusion_matrix(y_test, pred_dt)
          cm_display = metrics.ConfusionMatrixDisplay(confusion_matrix = confusion_matrix,
                                                      display_labels = [True, False])
          cm_display.plot()
          plt.show()
```

Figure 17

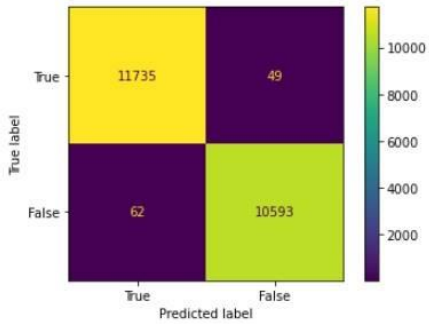

Figure 17a

- Confusion matrix of Random Forest Classifier

```
In [95]: confusion_matrix = metrics.confusion_matrix(y_test, pred_rfc)
         cm_display = metrics.ConfusionMatrixDisplay(confusion_matrix = confusion_matrix,
                                                     display_labels = [True, False])
         cm_display.plot()
         plt.show()
```

Figure 18

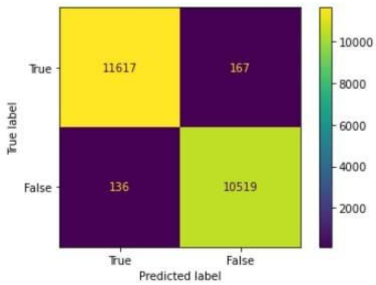

Figure 18a

- Confusion matrix of Gradient Boosting Classifier

```
In [100]: confusion_matrix = metrics.confusion_matrix(y_test, pred_gbc)
          cm_display = metrics.ConfusionMatrixDisplay(confusion_matrix = confusion_matrix,
                                                      display_labels = [True, False])
          cm_display.plot()
          plt.show()
```

Figure 19

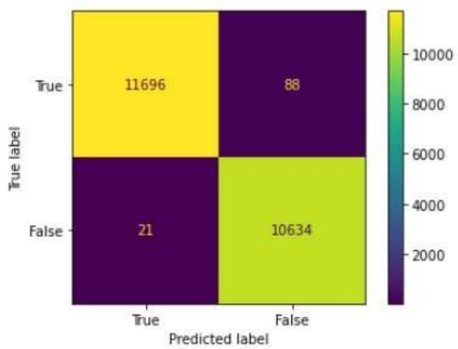

Figure 19a

L. We manually tested our models with the manual testing dataset we had saved. This is shown in the figure below.

Figure 20

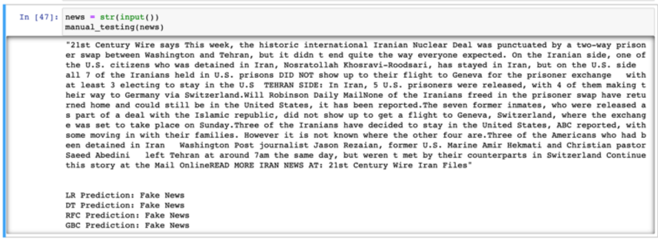

Figure 20a

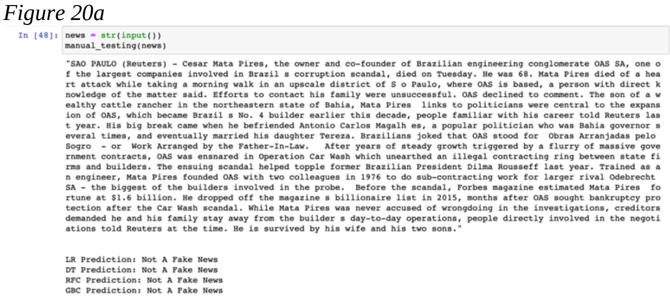

Figure 20b

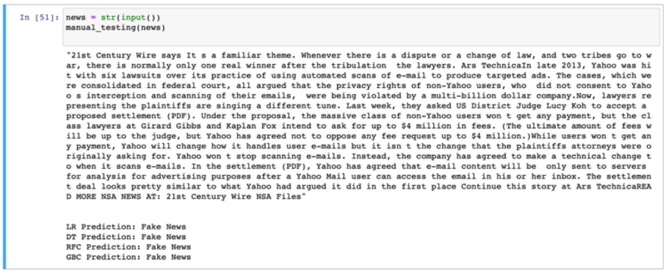

Figure 20c

M. METRIC COMPARISON

- **Comparison of Precision, Recall, F1-scores and Accuracy for all four classifiers- for predicted class '0.'**

classifiers	Precision	Recall	F1 score	Accuracy
Logistic Regression	0.99	0.99	0.99	0.99
Decision Tree	0.99	1.00	1.00	0.99
Random Forest	0.99	0.99	0.99	0.99
Gradient Boosting	1.00	0.99	1.00	1.00

Table 1:Comparison Table for Class 0

- **Comparison of Precision, Recall, F1-scores and Accuracy for all four classifiers- for predicted class '1'**

classifiers	Precision	Recall	F1 score	Accuracy
Logistic Regression	0.98	0.99	0.99	0.99
Decision Tree	1.00	0.99	0.99	0.99
Random Forest	0.99	0.99	0.99	0.99
Gradient Boosting	0.99	1.00	1.00	1.00

Table 1a:Comparison Table for Class 1

N. ACCURACY IN PERCENTAGE OF THE CLASSIFIERS

classifiers	Accuracy
Logistic Regression	99%
Decision Tree	99%
Random Forest	99%
Gradient Boosting	100%

Table 2:Accuracy Percentage

5. CONCLUSION AND FUTURE WORK

More people now prefer to read their news on social media platforms like Facebook, TikTok, Instagram, twitter, WhatsApp, and others rather than through traditional news sources, as evidenced by the recent rise in social media followers on these platforms. Due to these unprecedented followers, the social media has been targeted by many as a platform for spreading fake news which sometimes can be detrimental to both individual users and the society at large. Newspapers that were formerly favored as online news pieces and social media sites like Facebook, Instagram, TikTok, Twitter, WhatsApp etc., are gradually replacing printed copies. The spread of fake news only complicates matters and seeks to alter or impede people's attitudes and beliefs about using digital technologies. When someone is misled by false information, they may begin to think that what they know about a specific subject is inaccurate. We have therefore created a mechanism for detecting fake news that uses user input to determine if it is true or false to stop the epidemic. The usage of multiple NLP and machine learning techniques is required. An adequate dataset is used to train the model, and its performance is evaluated using various performance metrics.

The news headlines or articles are categorized using the best model with the highest accuracy score. After training and testing our models, the Gradient Boosting classifier turned out to be the best model with an accuracy score of 100% as shown in the analysis above. However, other classifiers are also good because they have an accuracy score of 99% each. Therefore, we can state that there are 99% to 100% possibilities that a user's input will be categorized to its true

nature if they submit a particular news article or its headline into our model.
For future work, we recommend using datasets from other repositories to analyze the model. Also, other feature extraction techniques, such as count vectorizer, can be used, and results can be compared with this study. Furthermore, different machine learning algorithms can be used to compare their performance to the existing ones.

The news headlines or articles are categorized using the best model, or the model with the highest accuracy. Our best model turned out to be Gradient Boosting with a 100% accuracy rate, as shown in the analysis above. However, other classifiers are also good because they have an accuracy score of 99% each. Therefore, we can state that there are 99% to 100% possibilities that a user's input will be categorized to its true nature if they submit a particular news article or its headline into our model. For future work, we recommend using datasets from other repositories to analyze the model. Also, other feature extraction techniques, such as count vectorizer, can be used, and results can be compared with this study. Furthermore, different machine learning algorithms can be used to compare their performance to the existing ones.

REFERENCES

[1] K. Agarwalla, S. Nandan, and D. D. Varun Anil Nair, 'Fake News Detection using Machine Learning and Natural Language Processing', *International Journal of Recent Technology and Engineering (IJRTE),* no. 7, 2019.

[2] I. Ahmad, M. Yousaf, S. Yousaf, and M. O. Ahmad, 'Fake News Detection using Machine Learning Ensemble Methods', *Complexity*, vol. 2020, pp. 1–11, 2020.

[3] *'Fake News: Real Lies Affecting Real People'*. North Charleston, SC: Createspace Independent Publishing Platform, 2018.

[4] S. Brown, 'Machine learning, explained', *MIT Sloan,* 21-Apr-2021. [Online]. Available: https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained. [Accessed: 09-Dec-2022].

[5] D. Chaffey, 'Global Social Media Statistics Research Summary 2022', *Smart Insights,* 22-Aug-2022. [Online]. Available: https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research. [Accessed: 09-Dec-2022].

[6] '3 Types Of Machine Learning You Should Know', *Coursera,* 14-Mar-2022. [Online]. Available: https://www.coursera.org/articles/types-of-machine-learning. [Accessed: 09-Dec-2022].

[7] Z. Khanam, B. N. Alwasel, H. Sirafi, and M. Rashid, 'Fake News Detection Using Machine Learning Approaches', *IOP Conf. Ser. Mater. Sci. Eng.,* vol. 1099, no. 1, p. 012040, 2021.

[8] U. Sharma and S. Saran, 'Fake News Detection using Machine Learning Algorithms', *International Journal of Engineering Research & Technology (IJERT),* vol. 9, no. 3, pp. 509–518, 2021.

[9] L. Waikhoma and R. S. Goswamib, 'Fake News Detection Using Machine Learning', in *International Conference on Advancements in Computing & Management,* 2019.

[10] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, 'Fake News Detection on Social Media: A Data Mining Perspective, *SIGKDD Explore.,* vol. 19, no. 1, pp. 22–36, 2017.

[11] 'DataReportal – global digital insights', *DataReportal – Global Digital Insights.* [Online]. Available: https://datareportal.com/. [Accessed: 11-Dec-2022].

[12] J. M. B, *Fake News: Real Lies, Affecting Real People*. North Charleston, SC: Createspace Independent Publishing Platform, 2018.

[13] 'International Journal of Science and Research (IJSR) ISSN: 2319-7064 ResearchGate Impact Factor (2018): 0.28 | SJIF', International Journal of Science and Research, pp. 7–426, 2018.

[14] H. Allcott and M. Gentzkow, 'Social Media and Fake News in the 2016 Election', *J. Econ. Perspect.*, vol. 31, no. 2, pp. 211–236, 2017.

[15] J. Niall, V. L. Conroy, and Y. Rubin, 'Automatic Deception Detection: Methods for Finding Fake News', *Proceedings of the Association for Information Science and Technology,* vol. 52, pp. 1–4, 2015.

[16] E. Mustafaraj and P. T. Metaxas, 'The Fake News Spreading Plague: Was It Preventable?', *arXiv [cs.SI],* S.I.17.

[17] R. S. Nickerson, 'Confirmation Bias: A Ubiquitous Phenomenon in Many Guises', *Review of General Psychology,* vol. 2, pp.175-220, 1998.

[18] A. Ward and T. Turiel, 'Naïve Realism in Everyday Life: Implications for Social Conflict and Misunderstanding', *Values and knowledge,* pp. 103–135, 1997.

[19] D. Kahneman and A. Tversky, 'Prospect Theory: An Analysis of Decision Under Risk. Econometrical', *Journal of the Econometric Society,* pp. 263–291, 1979.

[20] D. M. J. Lazer et al., ‘The Science of Fake News’, *Science*, vol. 359, no. 6380, pp. 1094–1096, 2018.

[21] B. Nyhan and J. Reifler, ‘When Corrections Fail: The Persistence of Political Misperceptions’, *Political. Behavior.*, vol. 32, no. 2, pp. 303–330, 2010.

[22] A. Tversky and D. Kahneman, ‘Advances in Prospect Theory: Cumulative Representation Of Uncertainty, *Journal of Risk and Uncertainty*, vol. 5, no. 4, pp. 297–323, 1992.

[23] H. Tajfel and C. John, ‘An Integrative Theory of Intergroup Conflict’, *The Social Psychology of Inter-Group Relations*, vol. 33, 1979.

[24] H. Tajfel and J. C. Turner, ‘The Social Identity Theory of Intergroup Behaviour’, In *Political Psychology*, Psychology Press, 2004, pp. 276–293.

[25] E. Solomon and H. Asch, ‘Effects of Group Pressure Upon the Modification and Distortion of Judgment in Groups, Leadership, And Men’, *Organizational Influence Processes, vol 58*, pp. 295-303, 1951.

[26] M. Gentzkow, J. M. Shapiro, and D. F. Stone, ‘Media bias in the marketplace: Theory’, in *Handbook of Media Economics*, Vol. 1, pp. 623-645, 2015.

[27] J.-N. Kapferer, Rumors: Uses, Interpretation and Necessity. 2017.

[28] E. Ferrara, O. Varol, C. Davis, F. Menczer, and A. Flammini, ‘The Rise of Social Bots’, *Communications of the ACM*, vol. 59, no. 7, pp. 96–104, 2016.

[29] A. Bessi and E. Ferrara, ‘Social Bots Distort the US Presidential Election Online Discussion’, *First Monday*, vol. 21, no. 11, 2016.

[30] J. Cheng, M. Bernstein, C. Danescu- Niculescu-Mizil, and J. Leskovec, ‘Anyone Can Become a Troll: Causes of Trolling Behaviour in Online Discussions’, in *CSCW* ‘17.

[31] Z. Chu, S. Gianvecchio, H. Wang, and S. Jajodia, ‘Detecting Automation of Twitter Ac- Counts: Are You a Human, Bot, Or Cyborg?’, *IEEE Transactions on Dependable and Secure Computing*, vol. 9, no. 6, pp. 811–824, 2012.

[32] D. Michela et al., ‘Echo Chambers: Emotional Contagion and Group Polarization On Facebook’, *Scientific Reports*, vol. 6, 2016.

[33] M. Del Vicario et al., ‘Echo Chambers in The Age of Misinformation’, *arXiv [cs.CY]*, C.Y.15.

[34] D. Michela et al., ‘The Spreading of Misinformation Online’, *Proceedings of the National Academy of Sciences*, vol. 113, pp. 554–559, 2016.

[35] B. Robert, ‘Attitudinal Effects of Mere Exposure’, *Journal of Personality and Social Psychology*, vol. 9, no. 2p2, 1968.

[36] C. Paul and M. Matthews, ‘The Russian ‘Firehose of Falsehood’ Propaganda Model: Why It Might Work and Options to Counter It,’ 2008.

[37] B. Robert, ‘Mere Exposure: A Gateway to The Subliminal’, *Current Directions In Psychological Science*, vol. 10, pp. 224–228, 2001.

[38] V. Rubin, N. Conroy, Y. Chen, and S. Cornwell, ‘Fake News or Truth? Using Satirical Cues to Detect Potentially Misleading News’, in *Proceedings of the Second Workshop on Computational Approaches to Deception Detection*, 2016.

[39] Researchgate.net. [Online]. Available: https://www.researchgate.net/publication/319255985_Automatic_Detection_of_Fake_News.

[40] R. Mihalcea and C. Strapparava, ‘The Lie Detector: Explorations in The Automatic Recognition of Deceptive Language’, in *Proceedings of the ACL-IJCNLP*, 2009.

[41] R. Feng, ‘Syntactic stylometry for deception detection’, in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers*, vol. 2, Association for Computational Linguistics, 2012, pp. 171–175.

[42] A. M. Kaplan and M. Haenlein, ‘Users of The World, Unite! The Challenges and Opportunities of Social Media, *Bus Horiz*, vol. 53, no. 1, pp. 59–68, Jan. 2010, doi: 10.1016/j.bushor.2009.09.003.

[43] K. Tanaka, ‘& ACM Digital Library, in Proceedings of the 3rd workshop on Information credibility on the web, ACM, 2009.

[44] A. M. Muniz Jr and T. C. O’Guinn, ‘Brand Community’, *J. Consum. Res.*, vol. 27, no. 4, pp. 412–432, 2001.

[45] R. V. Kozinets, ‘The Field Behind the Screen: Using Netnography for Marketing Research in Online Communities’, *J. Mark. Res.*, vol. 39, no. 1, pp. 61–72, 2002.

[46] L. Ramadhan, ‘TF-IDF simplified’, Towards Data Science, 20-Jan-2021. [Online]. Available: https://towardsdatascience.com/tf-idf-simplified-aba19d5f5530. [Accessed: 11-Dec-2022].

[47] Pushkar, ‘Different Techniques for Text Vectorization’ , *Codersarts* AI, 27-Jan-2021.

[48] A. Saini, ‘Gradient Boosting Algorithm: A Complete Guide For Beginners’, *Analytics Vidhya*, 20-Sep-2021. [Online]. Available: https://www.analyticsvidhya.com/blog/2021/09/gradient-boosting-algorithm-a-complete-guide-for-beginners/.

[49] S. R. Safavian and D. Landgrebe, ‘A Survey of Decision Tree Classifier Methodology’, *IEEE Trans. Syst. Man Cybern.*, vol. 21, no. 3, pp. 660–674, 1991.

[50] V. Kumar, ‘Evaluation Of Computationally Intelligent Techniques For Breast Cancer Diagnosis’, *Neural Computing and Applications*, vol. 33, no. 8, pp. 3195–3208, 2021.

[51] J. R. Quinlan, *Discovering Rules by Induction from Large Collections Of Examples. Expert Systems in The Microelectronics Age*. Edinburgh: Edinburgh University Press, 1979.

[52] ‘Basic Evaluation Measures from The Confusion Matrix’, *Classifier Evaluation with Imbalanced Datasets*, 03-Jun-2015. [Online]. Available: https://classeval.wordpress.com/introduction/basic-evaluation-measures/. [Accessed: 12-Dec-2022].

[53] ‘Association rule’, *GeeksforGeeks*, 13-Sep-2018. [Online]. Available: https://www.geeksforgeeks.org/association-rule/. [Accessed: 12-Dec-2022].

[54] A. Zubiaga, A. Aker, K. Bontcheva, M. Liakata, and R. Procter, *‘Detection and Resolution of Rumours in Social Media: A survey’*, 2017.

[55] L. Wu, J. Li, X. Hu, and H. Liu, ‘Gleaning Wisdom from The Past: Early Detection Of Emerging Rumours In Social Media’, [Online]. Available: http://bit.ly/2as5MJ5

[56] J. Sampson, F. Morstatter, L. Wu, and H. Liu, ‘Leveraging the Implicit Structure Within - Social Media for Emergent Rumour Detection’, in CIKM’15,

[57] Y. Li et al., ‘A Survey On Truth Discovery’, *SIGKDD Explore,* vol. 17, no. 2, pp. 1–16, 2016.

[58] S. Mukherjee and G. Weikum, ‘Lever- Aging Joint Interactions for Credibility Analysis in News Communities’, in CIKM’15.

[59] G. Weikum, ‘What Computers Should Know, Shouldn’t Know, And Shouldn’t Believe’, in *Proceedings of the 26th International Conference on World Wide Web Companion - WWW ’17 Companion*, 2017.

[60] P. Biyani, K. Tsioutsiouliklis, and J. Blackmer, ‘” 8 Amazing Secrets for Getting More Clicks”: Detecting Clickbaits in News Streams Using Article In- Formality’, in *AAAI’*,16.

[61] A. Chakraborty, B. Paranjape, S. Kakarla, and N. Ganguly, ‘Stop Clickbait: Detecting and Preventing Clickbaits In Online News Media’, in ASONAM, 16.

[62] J. N. Blom and K. R. Hansen, ‘Clickbait: Forward-Reference as Lure in Online News Headlines’, *Journal of Pragmatics,* vol. 76, pp. 87–100, 2015.

[63] M. Potthast, S. Köpsel, B. Stein, and M. Hagen, ‘Clickbait Detection’, in Lecture Notes in Computer Science, Cham: Springer International Publishing, 2016, pp. 810–817.

[64] U. Sharma, S. Saran, and S. M. Patil, “Fake News Detection using Machine Learning Algorithms.” [Online]. Available: *www.ijert.org*

[65] C. Dewey, *Facebook has repeatedly trended fake news since firing its human editors. The Washington Post*. 2016.

[66] *Kaggle.com*. [Online]. Available: https://www.kaggle.com/search. [Accessed: 14-Dec-2022].

[67] M. Dollarhide, ‘Social media: Definition, effects, and list of top apps’, Investopedia, 02-Nov-2010. [Online]. Available:https://www.investopedia.com/terms/s/social-media.asp.

[68] B. Auxier and M. Anderson, ‘Social media use in 2021’, *Pew Research Center: Internet, Science & Tech*, 07-Apr-2021. [Online]. Available: https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/.