The Role of AI in Combating Fake News and Misinformation

PREPARED BY

PARTHA SARATHI PURKAYASTHA (2020A7PS0043P)

NACHIKET KOTALWAR (2020A7PS0024P)

LABEEB AHSAN (2020A7PS0045P)

APRIL 2022

ABSTRACT

The recent proliferation of social media has undoubtedly brought about many benefits, but along with it also a serious impediment to the society in the form of "Fake News" which has become an eminent barrier to journalism, freedom of expression, and democracy, as a whole. Along with differentiating between different types of fake news and misinformation, the study also aims to understand the extent of social media as a source of information and the impact of AI-based detection techniques in combating them, with the help a questionnaire involving university students where we focus on the following dimensions: extent of social media as a source of news, news verification practices, and frequency of encountering fake news. We analyzed currently used techniques to detect fake news, identify their shortcomings and compare them with the emerging models. We compared the performances of the models in light of parameters like the words used, the words before and after them, and the overall relationships between the words in a text. We also compared the models on a dataset based on social media and compared the changes in performance using Ensemble Learning approaches. The study aimed to identify suitable models for Fake News Detection. This is in hopes to eventually promote a safe and healthy environment for sharing information and content online, and in the process, help develop strategies and techniques to curb the spread of fake news on social media.

Keywords: Fake news, Machine learning, Ensemble Learning, Artificial Intelligence, Semantics, Syntax, Algorithms, Social media

1. INTRODUCTION

Fake News Definition

The term "fake news" has been around for decades, but it has only recently gained traction in the media and among politicians. It describes any news story that someone believes is untrue because it has factual inaccuracies, was designed to mislead people, or is deliberately false. However, we find there is no accurate definition of the term "fake news." So, we follow the one that has been widely adopted in recent studies - Fake news is a news article that is intentionally and verifiably false. Thus, we consider a piece of article to be false news if the article includes false information that can be verified as such, and it is created with the dishonest intention of misleading consumers.

According to Claire Wardle (2017) the problem of information pollution can be divided into three main types of problems -

Mis-information - False news spread unintentionally

Dis-information - False news propagated knowingly

Mal-information - True news shared with a malicious intent

Because of its intent, disinformation can cause significant harm especially disinformation related to politics. Fake news about COVID-19 is an example of misinformation. Although Malinformation can be equally harmful, as it is true we don't classify it as fake.

Fake News : Propagation and Effects

Fake news has had a massive impact on today's world. It has led to people having to take precautionary steps both in their social and work lives. It has been seen that fake news has affected interpersonal relationships as well according to the study conducted by Duffy et al. (2019). Fake news can deteriorate the trust between people, with the sharers of fake news feeling guilty and embarrassed. Most of the time, it is in a rush to get fresh news to close ones that most people share fake news, which effectively leads to the same close ones doubting further

information from the spreader. In a few cases, the spreader was repeatedly asked to verify his source before anyone read the shared information. In research from Alonso García et al. (2020) it was found that fake news has also affected the scientific community. With exponential increase in the fake news being circulated around 2016, the number of scholarly articles on them saw a rise itself since around the same time and is still on the rise. Fake news has impacted society quite deeply by creating confusion about any available information that a new line of research can be seen emerging with the sole objective to carry out the in depth analysis of different aspects of fake news. Even large companies have not been unaffected by fake news according to Flostrand et al. (2019). Most of the companies have adopted certain precautionary measures to ensure that they themselves do not accidentally spread fake news and to ensure that they are less affected by viral fake stories about them. They concluded that fake stories about their organizations were more harmful than certain senior executives having a reputation for bad reasons. It was also said that fake news is there to stay and has become an integral part of a companies' image management. Employees have been affected as well (Lee et al., 2019). Employees of companies which were spreading fake news via advertisements were seen to not care about their companies' image to the extent that they would degrade its value in front of customers or external stakeholders. Such employees were observed to leave and change their companies as well. According to a study conducted by Rocha et al. (2021) people's health also took a toll due to the fake news being shared about COVID-19. It was found that social media was one of the main sources of fake news and conspiracy theories. These caused people to be skeptical of the rules imposed by the Governments, advice of researchers and health professionals. This also led to the worsening of the COVID-19 pandemic which further resulted in panic, anxiety, depression and such mental illness to people of all age groups. Fake news has the potential of even costing people lives as was the case where at least 20 people were killed in 2018 as a result of the circulation of a single piece of fake news on WhatsApp. (Vij, 2018).

Methods for Fake News Detection

To tackle the problem of classifying fake news, possible solutions are expert based fact-checking and crowd-sourced fact-checking. In expert based fact-checking, professional

fact-checkers are employed to detect and flag false content, however this approach is limited in scope due to high costs and the sheer volume of news content to fact-check.

Crowd-sourced fact-checking is based on the idea of wisdom of crowds. By this approach, entire publications/websites can be rated based on a large number of people thus with minimal costs. Although normal people are not the best judges of accuracy of new articles in many cases and according to Pennycook and Rand (2019), they judge around 40% of true news stories as false and 20% of false stories as true, and their personal biases can also affect the accuracy of identifying false articles.

An emerging method to tackle this problem is AI based classification. Artificial intelligence is a branch of Computer Science that aims to simulate human intelligence with the help of specialized hardware and software by writing, training, and implementing machine learning algorithms. Various applications of AI include tumor detection, natural language processing, language translation, speech recognition, and computer vision. Python, R, and Java are the popular programming languages used to currently implement the models in this field. We have used Python to implement our models. AI programming focuses on three primary cognitive skills: learning, reasoning, and self-correction.

- Learning- The main focus here is on collecting data and creating algorithms providing step-by-step instructions to turn raw data into something tangible and accomplish a specific task.
- Reasoning- The main focus here is on identifying the best algorithm to achieve the desired result.
- Self-correction- The main focus here is to continuously fine-tune and improve algorithms to provide the most accurate results possible.

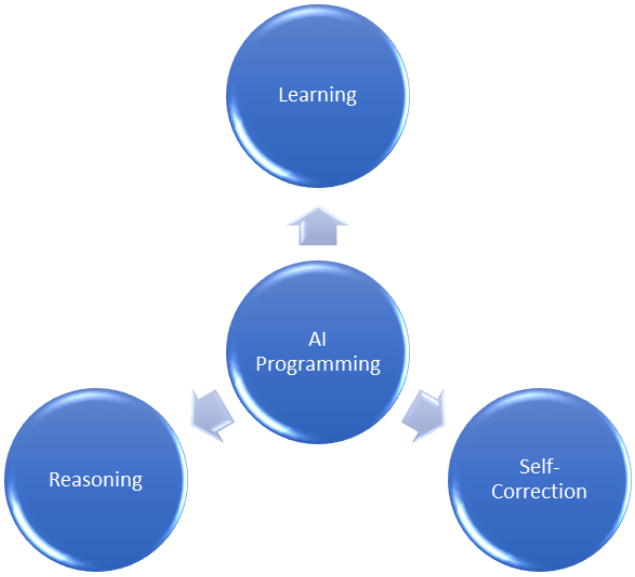


Figure 1: Three Primary Cognitive skills in AI programming

AI Models require large amounts of labeled training data, which are then analyzed via algorithms to form certain correlations between the inputs and outputs and make predictions about the future states. In this fashion, we aim to train a model that can learn to identify and classify fake news by reviewing millions of examples.

1.1. Rationale of the study

This study aims to analyze the impact of fake news on society and the current AI technologies used to detect fake news and contain misinformation. From a social perspective, we realized that social media is rapidly replacing traditional media as a source of news. People knowingly and unknowingly contribute to fake news since there are no authenticity checks before posting anything on such platforms. As a result, someone busy with their problems in everyday life will not take the time to verify every piece of news they receive and thus fall prey to misinformation. Furthermore, not everyone is always up to date on current events to be able to classify and disregard any piece of content as fake news. Thus, we did a study on the variables listed above in order to have a better understanding of the current social condition, and to get a quantitative measure of how much social media contributes as a source of news and more importantly, fake news.

From an AI point of view, we aim to train and compare the efficiencies of the basic models used to classify fake news, with new memory-based and Ensemble approaches. With the help of the results obtained, we intend to identify the various features of texts that play a crucial role in classifying news, which can further be incorporated into existing algorithms to build better and more accurate models.
By conducting a survey with the target demographic of 18-22 year-old Indian university Students , we are targeting the most involved generation in social media and possibly in the propagation of false information, along with their parents to compare the practices.

2. REVIEW OF LITERATURE

Fake News

1) Fake News and Social Media

Based on the study from Jonathan Albright (2017), only fact based evidence is not enough anymore for a significant chunk of the population. Social media, especially facebook has changed the paradigm in sharing news. The study the challenges of trust, transparency, and reducing attention span due to curated feeds to every user developed through intimate details of the person. It notes the need to collect more data about how fake news reaches and impacts the consumers.

Al-Zaman (2021) analyzed misinformation originating from 138 different countries. They found that India, USA, Brazil and Spain are the countries most affected with misinformation. The internet alone accounts for 90.5% of the amount of misinformation out of which social media is responsible for 84.94% of the entire amount of misinformation. Due to India's higher internet penetration, social media usage and lower digital literacy, India is the worst hit country due to COVID-Misinformation. Facebook alone accounted for over two thirds (66.87%) of misinformation across platforms.

The study by BALI and DESAI (2019) dives into the details of fake news and its effects in overall and then the Indian context. Instances like the Cambridge Analytica Scandal - (A political consulting firm, Cambridge Analytica illegally used data from 87 million facebook accounts to influence the outcome of the US Presidential elections.) and others are discussed. Fake news has become a great threat to society and the framework of democracy. Social media has become a major player in not only propagation but also generation of information. In the Indian context, WhatsApp has been the major driver of fake news, triggering violence in extreme cases. The study includes various cases of lynching and violence which occured in India due to fake news. Most of the mechanisms to stop the spread of fake news are self-regulatory, mostly arising out of criticism. It identifies the main problem as not the technology but the way we use it. Some of the solutions like policy intervention, participation by investigatory agencies, and more importantly to increase awareness are proposed and analyzed.

2) Propagation

Duffy et al.(2019) carried out a study to gauge the impact of sharing fake news on interpersonal relations. A survey was conducted where 12 groups of 88 Singaporean adults were made with the groups being made age wise, to group together similar news consumption and sharing habits. These groups were asked about their reactions to fake news and their sources of news. The conclusion stated that most people share fake news thinking it's real to strengthen their relations with other people. Once a person knows he has shared fake news to someone he is likely to be more cautious of the news he is sharing due to the negative reactions of other people. It was also discovered that there was a difference in the motivations for sharing across generations. They found that "Older participants who wished to be seen as sources of advice and wisdom – opinion leaders – were less critical of the stories they shared, while younger participants were more circumspect." (p. 11). It was also noticed that people tend not to check news backing their beliefs even if they might seem like fake news.

A study by Nagi (2018) focused on the impact of new social media in generating fake news and in turn impact of fake news on the society. The research examined contemporary challenges using typical empirical-analytical methods based on the little data accessible in open access

repositories about false news. The author's findings and comments are based on data from the Pew Research Center (USA), the Reuters Institute (UK), the European Commission (EC), and other sources, as well as the author's own poll performed in an Executive MBA (EMBA) programme at National Economic University in Hanoi, Vietnam. The results stated that fake news has become a problem with the origination of these new social media platforms as supported by the fact that the term ‘fake news’ was barely searched on google before 2016. However, according to a Pew Research Center survey conducted in 2016 39% of the population are very confident, 45% are somewhat confident, 9% are not very confident, and 6% are not confident at all. Nagi (2019) also found that ‘About a quarter (23%) say they have never shared such stories, while roughly equal proportion say that they have shared made-up news knowingly and unknowingly’(p. 86). It was discovered that poor and wealthy countries are both as likely to use social media as their source of news.

Tandoc et al. (2019b) carried out a study with the objective to understand how disinformation diffuses through social media and analyse how and why social media users respond to fake news. It used a mixed-methods approach in an explanatory-sequential design. It combines results from in-depth interviews with 20 participants from Singapore and a national survey consisting of 2051 respondents also from Singapore. It was found that upon knowing that a certain news was fake news on social media, the users would correct it only if they had connections with a very serious issue or had some personal significance to them, otherwise, they would ignore the correction.

3) Impact

Work by Flostrand et al.(2019) consisted of a Delphi study of a panel consisting of 42 academics who have peer-reviewed publications in the brand management domain. According to the panel fake news was there to stay and has forever become a part of the brand’s image management. When asked to compare the impact of fake news to other negative organizational events, the panel ranked viral fake news stories that deteriorate the brand's values as more harmful than having senior executives having a bad reputation for misdeeds. However, data breach was still more damaging to a brand’s value compared to viral fake news. Other than companies getting affected by viral fake news, their employees were affected by fake news spreaded by their companies as well. Lee et al. (2019) conducted a study that looks at what effects fake company

slogans have on their employees. It was carried out using secondary data on topics like fake news, employee audiences, andcompany slogans. It was found that employees pay attention to their company slogans and having fake slogans would definitely deteriorate the company's standing in public. Employees of companies with fake slogans were found to degrade their company in front of customers and external stakeholders, many would leave the companies and look for alternatives as well.

Alonso García et al. (2020) carried out a study with the objective to analyze the impact of fake news on the research community. This work advocated a 'scientometric-type methodology, through scientometric laws, impact indicators, and scientific evolution of 640 publications of the web of science (WOS)'(p. 1). It was around 2016 when fake news started to increase. In 2017 when there was exponential growth of fake news, articles on them in the web of sciences database increased, from 57 in 2017 to 215 in 2018. A slight increase was observed from 2018 to 2019 as well. The rise of fake news has had quite an influence on society and within the research community, creating confusion about any information on the internet. Hence, creating a new line of research can be observed to study this phenomenon of fake news in depth.

Even people's health was deteriorating with fake news and conspiracy theories being circulated according to the study carried out by Rocha et al (2021) which focused on analyzing the impact of fake news during COVID-19 on people's health and social media. The findings reported that during the time of the COVID-19 pandemic social media was the major contributor in the spreading of fake news. With information about the pandemic becoming a part of people's life this caused issues like distrust in Governments, researchers, health professionals, which could have a significant impact on their health. Panic, depression, fear, psychological distress were among many other health problems that people had because of the spread of fake news irrespective of their ages.

Also based on research from Apuke & Omar (2021), misinformation propagated on social media has caused public anxiety with respect to the coronavirus. The study is based on a sample of 385 Nigerians, regarding spreading of COVID related fake news. False, medically unproven claims were widely circulating amongst the population. They developed a predictive model to analyze based on the factors of altruism, entertainment, socialization, pass time, and information sharing

and seeking. Altruism was the strongest factor followed by the tendency to share information which was also concluded by Duffy et al.(2019). It concluded that social media users' motivation for information sharing and seeking, and socialization correctly predicted the sharing of false information. However, with respect to the motivation of entertainment, no conclusive correlation was found.

4) Strategies to counter fake-news

The paper by Pennycook and Rand (2019) quantifies the effectiveness of crowdsourcing the problem of identifying fake news and compares the accuracy of the masses with the accuracy of experts. They conducted two studies with about 2000 participants, all of whom were US citizens. They found that common people across political opinions trusted mainstream media sources rather than politically biased or fake news sources. Hence credibility ratings obtained from crowdsourcing can be used as an important piece of data in algorithms.

Tagging social media articles with flags or labels has been identified as one of the important strategies. In a study with 717 participants, research from Gaozhao (2020) found that people relied heavily on the flags identifying a post as fake irrespective of their political background. Users are not good at differentiating misinformation from genuine information, due to lazy reasoning - reluctance to think critically, and motivated reasoning - thinking on the bases of preconceptions, confirmation bias and conservatism. It also found that the flags used to tag fake news on the internet are highly influential, irrespective of them being expert based or crowd-sourced. They don't promote thinking, rather act as a powerful influence. Hence the accuracy of applying flags is of paramount importance in solving the problem.

The article by Bernard Marr (2020) elaborates on the steps taken by the social media companies to combat COVID misinformation. Rumors and false news spread faster than before, especially in the backdrop of fear and anxiety. During the early stages of the coronavirus crisis, Twitter showed reliable sources like national agencies and the WHO on user searches. In the later stages, it relied on machine learning algorithms to flag content for removal, and also actively looked for commonly spread lies. Instagram and Facebook also employed algorithms to look for commonly used hashtags in fake news posts.

A study conducted by Khan et al.(2019) discusses false news and social media, with the goal of examining the relationship between the two, as well as how fake news is becoming a serious threat to civil society. Secondary data from many research papers relating to fake news and social media was analyzed to get the conclusion. The problem of fake news spreading on social media is a very grave problem. It is known to have caused multiple deaths in many countries and has caused violent protests and riots. It has the potential to grow a lot more with the increasing number of social media users. All of this is accelerated with the advancement of technology. Similar findings that inferred that social media was a major cause for the spread of fake news were found by Alonso García et al.(2020), Egelhofer and Lecheler (2019) and Nagi (2018). Detection and deletion of fake news using machine learning algorithms was a common proposed solution to prevent the current situation from deteriorating further.

Role of AI

We, as humans, are prone to bias. It might be due to us being reluctant to believe anything that goes against our beliefs, maybe due to everyone around accepting something as a fact or maybe if someone we look upto also agrees with something, thereby making us unreliable detectors of fake news. So we need some automated detection models to help us distinguish them. Although not always accurate, Machine learning techniques in AI have shown to help predict if a certain information is true or not with the help of patterns obtained from training data to a great extent. The different aspects of involvement of AI are described below:

1) Linguistic Characteristics

Linguistic indicators can help distinguish fake news from real ones. Natural Language Toolkit (NLTK) can be used to tokenize text, and count the occurrences of certain types of words, with averages per article of each type.

According to the findings of the study done by Rashkin H. et al.(2017), fake news articles generally are subjective, use pronouns in the first and second persons, and tend to exaggerate significantly. Reliable sources, however, use evidence based phrases like facts, figures, numbers, comparisons, and assertiveness.

2) Traditional Models

Traditional Machine Learning Algorithms such as simple SVMs (Support Vector Machines), Naïve Bayes, and Simple Neural Networks have shown to produce decent results. Aphiwongsophon et al. (2018) used SVM and Neural Network algorithms to analyze data from Twitter. Their work showed a 99.90 percent accuracy rate in detecting fake news.

However, they are not very consistent. They are often outperformed by Models using other forms of learning methods like LSTMs,Bi-LSTMs and Ensemble Learning Methods.(Ahmad et al., 2020;Vijayaraghavan et al.,2020).

Studies from Aldwairi and Alwahedi(2018) and Perez-Rosas et al.(2020) suggest that having larger training datasets may help enhance the classification performances of models as it gives them a bigger sample space to work upon.

3) Natural Language Processing

Natural language processing (NLP) approaches combined with machine learning algorithms process news information by detecting language patterns, sentiment, and occurrences of words commonly used in fake news.

Liu & Wu (2020) carried out a study proposing a Deep learning framework for Fake News Early Detection (FNED).The model gave promising results and worked in a content independent manner which could potentially be utilized to detect Deepfakes.

Vijayraghavan et al. (2020) attempted fake news detection as a binary classification problem and implemented different Natural Language Processing models such as TFIDF, CountVectorizer and Word2Vec models. These models are able to preserve most of the contextual information about

the text used and accurately detect fake news. In their work, a combination of CountVectorizer with LSTM achieved the best performance.

Along with CNNs and LSTMs, Sastrawan et al.(2021) also implemented Bidirectional LSTMs, combined with pre-trained word embedding. It was observed that Bi-directional LSTM-RNN model was significantly more effective than unidirectional models, outperforming them in multiple datasets. In the study done by Bahad et al.(2019) , which also used Bi-Directional LSTMs to predict fake news articles, it was observed that the Bi-LSTM model had less loss and more accuracy as compared to the Unidirectional LSTM model, although the differences in their performances was seen to be minimal.

4)Mixed approaches

The study conducted by Albahar et al.(2021) proposed a hybrid model based on an RNN and a SVM to detect rumors in news content. The authors used Fake news datasets of two subparts- the PolitiFact and GossipCop datasets. In both PolitiFact and GossipCop datasets, The proposed hybrid model was observed to have a greater Accuracy and F1 Scores than the other models considered in both the datasets considered.

Ahmad et al.(2020) proposed Ensemble learning methods for identifying patterns in text that help differentiate fake articles from true news. The dataset used was obtained from the World Wide Web, and contained news articles from all domains rather than just politics. They extracted different textual features from the texts and used them to train models. In the many datasets that were analyzed, Ensemble learners always remained above the average of individual learners in terms of accuracy.

2.1 Summary of Literature

Fake news mainly consists of misinformation and disinformation. Social media is one of the main modes of propagation of fake news. Fake news has affected how companies make their policies, how people think about each other. It has also affected people's health, with it helping in causing panic and confusion during the time of the COVID-19 pandemic. It has already caused

immense damage, because of which it is being studied everywhere in an attempt to stop the situation from escalating further. Some of the possible ways to reduce the spread of fake news include expert-based fact-checking and crowdsourcing. However, due to the magnitude of the problem, detecting and flagging fake news using machine learning algorithms was a commonly proposed solution.

Regarding the effectiveness of AI, not much can be concluded that applies to all models. Still, a common trend can be observed that if we introduce more aspects to texts such as memory, context, and sentience with the help of Neural networks and Recurrent Neural Networks, the overall effectiveness of the new models in comparison to the pre-existing models turns out to be greater than before. Most studies have the drawback that they do not have sufficiently large and verified datasets to train models properly, so this problem needs to be tackled. Currently, Bidirectional LSTMs seem much more promising than Unidirectional LSTMs- Unidirectional LSTM stores information of the past, but bidirectional LSTM manages inputs both ways- from past to future and from future to past. Ensemble Learning approaches have been show to perform quite efficiently as well, with future hopes of training hybrid models with better performances.

2.2. Research Gaps

In most of the research papers analyzed, it has been observed that they are not demographic-specific. Instead, the studies tend to focus on the population of mixed groups and then generalize the average obtained for all, thereby skewing the average greatly in the process. This method fails to consider aspects of the text that may have been specific to certain demographics - A significant difference can be observed in the styles of writing used by people from age groups 18-22 years and those from 30-55 years. As a result, when analysing new tweets from Twitter or old Facebook posts, a model may fail to correctly classify them as authentic or fake. Furthermore, most of the studies carried out are based in the US or the UK. So, we are focusing on an Indian audience. We are also collecting data targeting specific demographic groups - in our case, second and third-year university students in India along with their parents and relatives above 30 years of age, to bridge this gap and get an overall idea of the features

which would be necessary for future algorithms and build efficient models. The data obtained on news habits will provide insights on the potential impact of these models.

Concerning AI Models, since only a limited options for properly verified datasets are available online and those available are not large enough to train models accurately, most of the models reported under researches are very limited in their scope and application. So the results obtained from research and studies conducted cannot directly be generalized or extended to the pre-existing or emerging models. Larger datasets could help take a step forward in overcoming this gap. Moreover, context based datasets of articles could be a great step in bridging these gaps.

3. RESEARCH QUESTIONS AND OBJECTIVES

Null Hypothesis(H_0) - Memory-based learning approaches do not improve the accuracy of detecting fake news propagated on social media.

Alternative Hypothesis(H_1) - Memory-based learning approaches improve the accuracy of detecting fake news propagated on social media.

Objectives-

The two main objectives of this study are -

Objective 1: To analyze sources of news and the contribution of social media in fake news
Objective 2: To analyze effectiveness of AI in classifying fake news on a social media dataset, and identify important aspects of models

4. METHODOLOGY

4.1. Research Design

For the collection of primary data, we collected questionnaires specifically targeting two social groups - University students in 18-22 years of age and another targeting those above 30 years of age. The questions were designed to get an idea of the overall news consumption and verification practices of the two age groups.

We circulated the questionnaire mainly through WhatsApp groups. The questionnaires were sent in groups of varied branches in the university, and also forwarded to family WhatsApp groups. The email and the responses were kept anonymous so as to encourage the participants to tell the truth about their practices irrespective of the ethical or moral implications of their response.

Along with this, we are using a secondary dataset collected by Verma et al. (2021) wherein they designed the WELFake dataset by merging four popular news data sets (Kaggle,McIntire, Reuters, and BuzzFeed Political) and prepared a dataset of 72,134 news articles with 35,028 real and 37,106 fake news. The larger dataset was designed to prevent over-fitting of classifiers on a smaller dataset and enable better ML training.

The training data set used has five features:

Feature	**Description**
ID	Unique ID for a News Article
Title	The Title of a News Article
Text	The Text of the Article (can be incomplete)

Label	Indicates Reliability; **0** for unreliable **1** for reliable

We focussed mainly on the text features and implemented word embedding techniques to best classify news. Also, in our study we have changed the labels as 0 for reliable and 1 for unreliable news.

We have carried out fake news detection using five basic AI models and analyzed their effectiveness in successfully identifying fake news. The models used are Naive Bayes, Neural Networks using Keras, Support Vector Machines(SVMs), and Long Short-Term Memory(LSTMs). We then used Ensemble Approaches wherein we implemented Regression Forest Models, Memory based Ensemble Model and a Non-Memory based Ensemble Model in an attempt to build efficient classifying models.

4.2. Population

Analyzing from a social perspective, the population consists of university students, more specifically engineering students studying in India, and parents of students, with age above 30.

4.3. Sample

A total of 74 respondents participated in the study.

Age Demographic (Years Old)	Respondents
18-22 (Students)	51
30+ (Adults)	23

Students - Sophomores and Juniors pursuing their bachelors degree in engineering at the Birla Institute of Technology and Science - Pilani Campus across various branches of engineering. The questionnaire was sent across WhatsApp groups from which we got 51 responses in a span of 3 weeks. The age of the participants is between 18 and 22 years, the portion of the population which is significantly involved in the usage of social media.

Adults - Parents and elder relatives of the students over 30 years of age. Questionnaires were sent in WhatsApp groups. We got 23 responses over 3 weeks.

4.4. Sampling techniques

The questionnaire was only sent to select social media groups of second and third year students to ensure the demographic. Furthermore, it was sent in groups homogeneous in branches of study of students to make the sample less biased towards students pursuing a specific field of study. For the 30+ year age old demographic, the questionnaire was circulated in the parents' social media groups.

4.5. Measure Design

For the analysis of the models, we considered Accuracy, Precision, recall, F1-score and Specificity as the units of measurement. This measurement is done on a ratio scale and thus the values can be compared in terms of their absolute values.

Based on what proportion of outcomes the models provide are True Positives, True Negatives, False Positives and False Negatives, the different units of measurement are defined as below:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$precision = \frac{TP}{TP + FP}$$

$$recall = \frac{TP}{TP + FN}$$

$$F - score = \frac{2}{1/precision + 1/recall}$$

$$specificity = \frac{TN}{TN + FP}$$

True Positives(TP): Cases when a true outcome is labeled as true by the model
True Negatives(TN): Cases when a false outcome is labeled as false by the model
False Positives(FP): Cases when a false outcome is labeled as true by the model
False Negatives(FN): Cases when a true outcome is labeled as false by the model

- Accuracy is given by the ratio of the number of correct outputs (TP + TN), to the total number of samples. In cases of random datasets, judging models based on only this parameter may lead to wrong conclusions.
- Precision is given by the ratio between the number of samples correctly classified as fakes (TP), by the total set of positives predictions (TP + FP);
- Recall, also known as Sensitivity, is given as the ratio of the number of samples correctly classified as fakes (TP) and the total of samples that actually were fakes (TP + FN).
- Specificity is given as the ratio of the number of samples correctly classified as real (TN) and the total of samples that actually were real (TN + FP).
- F1-Score relates precision and Recall by a harmonic mean and the higher the value of the F1-Score, the better the classification.

To measure these parameters, we used the sklearn, keras, SVM and nltk libraries of Python language on the platform Google Collab, and the matplotlib library of Python to visualize

the Accuracies obtained as a Confusion matrix. A confusion matrix is a table used to describe the performance of a classification model on a set of test data for which the true values are known.

The questionnaire to collect data about news consumption and verification practices and the extent of use of social media, the questionnaire consisted of dichotomous and Likert scale based questions along with some choice based and short answer questions.

4.6. Dimensions of Research:

i. Dimensions under News -

a. Sources - Compare the various sources of news among people in general
b. Verification practices- Analyze how many people actually tend to verify the news they receive
c. Frequency - Analyze how often people come across fake news in their day-to-day life

ii. Dimensions under AI -

a. Basis of Classification - Analyze the features of text which are important in fake news detection
b. Accuracy-Measures if the predictions are correct on an average.
c. F1-Score- Also used for determining a model's accuracy on a dataset- a better statistic when there are imbalanced classes, as is in our case.
d. Precision- Measures how close the predictions are to one another.
e. Recall- Indicates how well our model can detect **relevant** data.
f. Specificity- Measures how correctly the models predict a negative result for cases which are actually negative, i.e. a True Negative rate

4.7. Analysis

For the primary data, we analyzed the mean data to get a quantitative idea of the people's behavior and differences and similarities between the two demographic age groups.We used Accuracy as the main parameter to gauge the performance of the various models implemented

against the fake news dataset. We also represented the Accuracy of the models in the form of a Confusion Matrix, which shows to what extent news was labelled correctly or incorrectly.

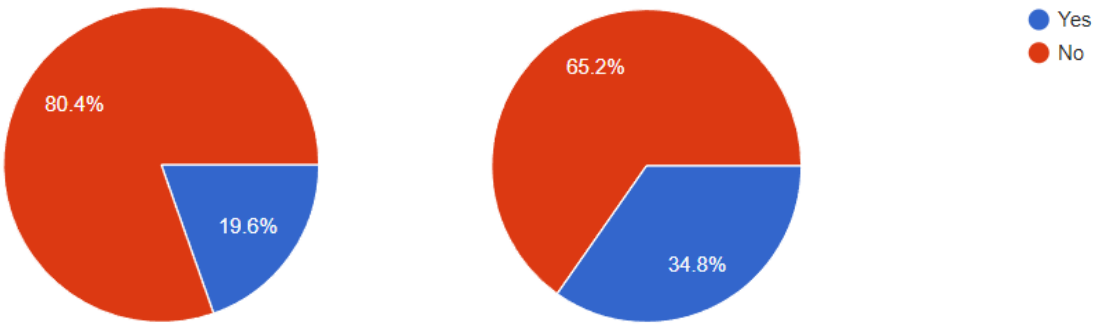
5. RESULTS AND DISCUSSION

5.1. Objective 1 :

To analyse sources of news and the contribution of social media in fake news.

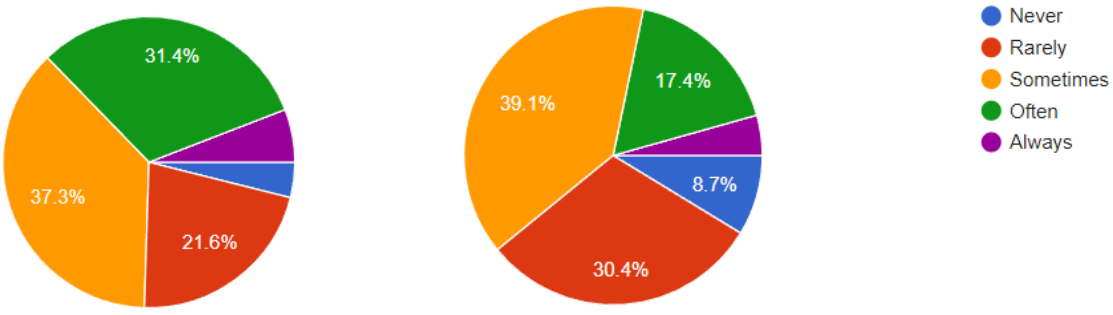
5.1.1. Analysis:

All the results obtained from the questionnaires collected have been attached in the appendix. Out of those, the key observations have been stated here in this section.



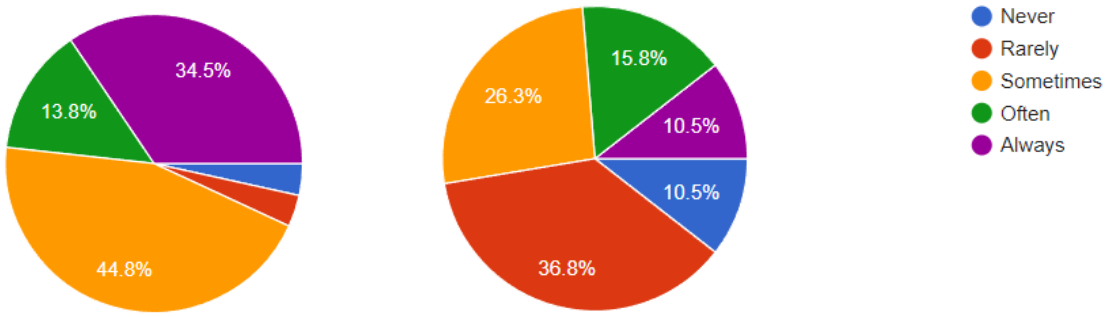
a) Students b) Adults

Figure 2 : News verification practices



a) Students b) Adults

Figure 3 : Frequency of Fake News



a) Students b) Adults

Figure 4 : Fake News Correction Practices

a. Source of News - The University students group consider Social Media along with News and other Mobile applications as their primary source of information while those above 30 cited Social media and Television as their primary sources.

Over 88% of the students listed social media as one of their primary sources. 65.2% of the adults also considered social media as one of their primary sources of news and information.

b. Social Media Sites - Among the Social media platforms used, the University students preferred YouTube, Instagram and WhatsApp as their 3 most used sources for news, while the above 30 groups used WhatsApp, YouTube and Instagram.

c. Subscriptions - Regarding Digital Subscriptions, 90.2% of those in the University Students group reported not paying for any digital subscriptions while 39.1% of those in the above 30 group reported subscribing so.

d. News Verification Practices - It was observed that over 80% of those in the University Students group did not use any fact-verification websites. A similar trend was observed in the Adult's group where this proportion was over 65% (Figure 2).

e. Frequency of Fake news - 37.3% of students reported to come across fake news sometimes, and 31.4% of them come across it often, while 2.9% of students reported that they never came across fake news.

39.1% of adults reported to come across fake news sometimes and 17.4% come across it often, while 30.4% adults reported to come across fake news rarely (Figure 3).

f. False News Correction Practices - Regarding the tendency to correct themselves after realizing they had spread something false, 36.8% of Adults reported that they rarely corrected themselves while 10.5% reported that they never corrected themselves afterwards. Only 10.5% always went ahead to correct themselves. This number is quite better in the case of Students where 34.5% said that they always corrected themselves and only 3.4% reported that they never corrected themselves (Figure 4). Tandoc et al. (2019b) also found that users would ignore the correction in news unless it was a very serious issue or had some personal significance.

g. Summary statistics for Numeric Questions - Our Questionnaire had 5 Numeric questions where response choices had been created based on a 1-5 Likert Scale.

Every question had one of the following two sets A and B, assigned the corresponding values-
Set A : 1-Never , 2-Rarely , 3-Sometimes , 4-Often , 5-Always
Set B : 1-Strongly Disagree , 2-Disagree , 3-Neutral , 4-Agree , 5-Strongly Agree

Questions (Abbreviated)	Mean	
	Students	Adults
Q1. How frequently do you come across fake news?(Set A)	3.13	2.78
Q2.How frequently do you try to verify the news you get from sources other than official channels?(Set A)	3.33	2.96
Q3.Technology and social media have made me a smarter and more informed person. (Set B)	3.88	2.91
Q4.The lack of control and fact-checking on social media makes it suitable for the propagation of unconfirmed and/or incorrect information.(Set B)	3.90	3.17
Q5.Traditional news outlets don't report fake news.(Set B)	2.51	2.83

5.1.2. Trends in Social media penetration in India

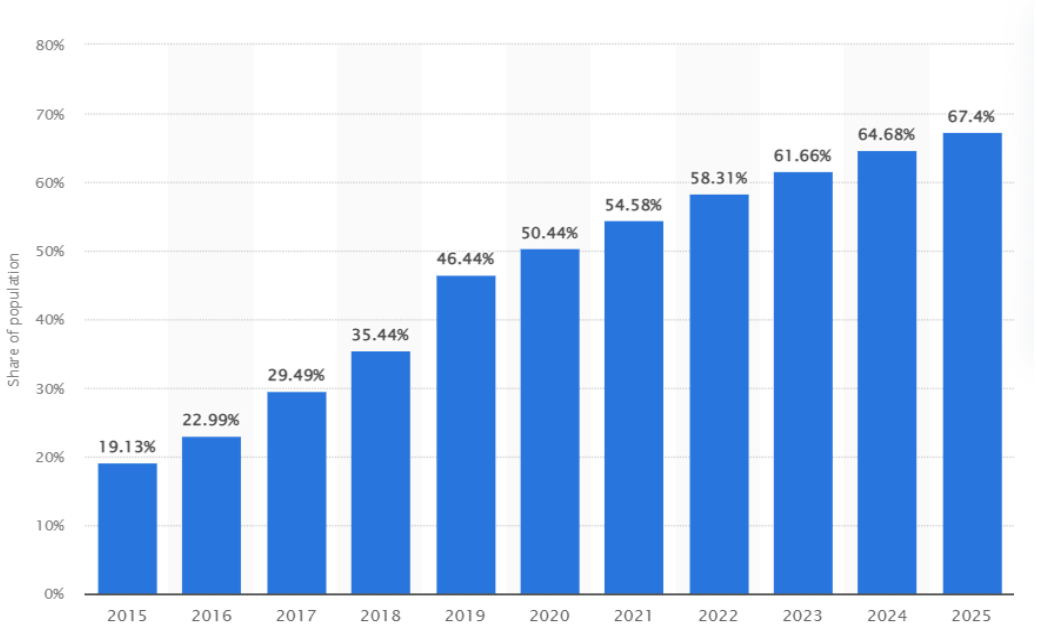


Figure 5 : Social network user penetration in India from 2015 to 2020, with estimates 2025[1]

The social media penetration of Indians has been increasing steadily leading to an increase in the fake news being circulated (Nagi, K. N. ,2018). This was concluded on the basis that the term 'fake news' was barely searched on google search engine before 2015. It was with the increase in the number of social media platforms and users in around 2016 that fake news circulation started increasing at such a rapid pace.

5.2. Objective 2: To analyze effectiveness of AI in classifying fake news on a social media dataset, and identify important aspects of models

5.2.1. Methodology:

The basic workings of the underlying algorithms are shown below (Dasaradh, 2020) :

[1] https://www.statista.com/statistics/240960/share-of-indian-population-using-social-networks/

a. Naive Bayes Classifier - Naive Bayes is a conditional probability model based on Bayes' theorem, which states that the conditional probability of an event Ck, provided x has already occurred, is given by -

$$p(C_k \mid \mathbf{x}) = \frac{p(C_k)\, p(\mathbf{x} \mid C_k)}{p(\mathbf{x})}$$

For each of the K possible outcomes, using the chain rule for repeated applications of the definition of conditional probability, this becomes -

$$\begin{aligned} p(C_k, x_1, \ldots, x_n) &= p(x_1, \ldots, x_n, C_k) \\ &= p(x_1 \mid x_2, \ldots, x_n, C_k)\, p(x_2, \ldots, x_n, C_k) \\ &= p(x_1 \mid x_2, \ldots, x_n, C_k)\, p(x_2 \mid x_3, \ldots, x_n, C_k)\, p(x_3, \ldots, x_n, C_k) \\ &= \cdots \\ &= p(x_1 \mid x_2, \ldots, x_n, C_k)\, p(x_2 \mid x_3, \ldots, x_n, C_k) \cdots p(x_{n-1} \mid x_n, C_k)\, p(x_n \mid C_k)\, p(C_k) \end{aligned}$$

In Naive Bayes, we assume that all the features x are mutually independent, and thus the probability now becomes -

$$p(C_k \mid x_1, \ldots, x_n) = \frac{1}{Z} p(C_k) \prod_{i=1}^{n} p(x_i \mid C_k)$$

This formula forms the algorithm for the Gaussian Naive Bayes classifier- which we used to train the model on the basis of the occurrences of a word in an article.

b. Support vector machine (SVM) - Support vector machine (SVM) are models generally considered as one of the most powerful classification models and are available in various kernels functions. The objective of an SVM model is to estimate a hyperplane (or decision boundary) on the basis of features set to classify data points. The dimensions of the hyperplane vary according to the number of features. As there could be multiple possibilities for a hyperplane to exist in an N-dimensional space, the task is to identify the plane that separates the data points of two classes with maximum margin. Mathematically, the cost function for the SVM model is given as:

$$J(\theta) = \frac{1}{2}\sum_{j=1}^{n}\theta_j^2,$$

And shown as such:

$$\theta^T x^{(i)} \geq 1, y^{(i)} = 1, \quad \theta^T x^{(i)} \leq -1, y^{(i)} = 0.$$

The function here uses a linear kernel. Kernels are used to build models when data points cannot be easily separated from one another, or are multidimensional in nature. We have used the Radial Basis Function(rbf) kernel to train our model here.

c. Neural Networks

Perceptrons are the simplest neural network that consists of n number of inputs, one neuron, and one output, where n is the number of features of our dataset. It consists of two main processes: Forward Propagation-The process of passing the data through the neural network, and the Learning Process-the way the data is used to train the model.

In Forward Propagation, we assign each input x_i with a weight w_i-which represents the strength of connection between the neurons, and calculate the dot product(sum of multiplied values). We then add a bias **b** to the dot-product to move the entire activation function left or right, so as to fit the model and generate the required output. Thus we get a final expression z as:

$$z = x.w + b$$

So as to introduce non-linearity, we now pass z through an Activation function- the most common being the Sigmoid function, which returns values in the range [0–1] and is particularly useful in Binary classification problems. The sigmoid function is given by:

$\hat{y} = \sigma(z) = \frac{1}{1+e^{-z}}$, where **σ** denotes the s*igmoid* activation function, and the output we get after the forward prorogation is known as the *predicted value* **ŷ**.

The learning algorithm consists of two parts: backpropagation- the algorithm for computing the gradient of the loss function with respect to the weights and bias, and

optimization- the task of assigning the best values for the weights and bias,upon which the model works most efficiently.

In Backpropagation, we calculate The Loss function (in our case the mean squared error,which squares the difference between actual (y_i) and predicted value ($\hat{y}_i$) to check how far we are from the desired solution) for the entire training dataset and term their average Cost function C, we get:

$$\frac{\partial C}{\partial w_i} = \frac{2}{n} \times sum(y - \hat{y}) \times \sigma(z) \times (1 - \sigma(z)) \times x_i$$

$$\frac{\partial C}{\partial b} = \frac{2}{n} \times sum(y - \hat{y}) \times \sigma(z) \times (1 - \sigma(z))$$

For optimization purposes we chose the gradient descent algorithm,which changes the weights and bias proportional to the negative of the gradient of the cost function with respect to the corresponding weight or bias. The Learning rate (α) is a hyperparameter which is used to control how much the weights and bias are changed at every step.
The weights and bias are updated as shown below, and the process of backpropagation and gradient descent is repeated until convergence, which is when the best values are determined.

$$w_i = w_i - \left(\alpha \times \frac{\partial C}{\partial w_i}\right)$$

$$b = b - \left(\alpha \times \frac{\partial C}{\partial b}\right)$$

This is the working of a single neuron but can be extended to entire neural networks with some modifications at key steps.

We have used Keras for our implementation of Neural Networks.

d. Long Short Term Memory Networks (LSTMs) - If we have a string of words in a sentence, every word has some relationship with another and this is very important when classifying articles.Traditional neural networks, however, cannot store memories of previous events to influence the later ones. Recurrent neural networks(RNNs), which are networks with loops allowing information to persist, address this issue.

Long Short Term Memory networks(LSTMs) are a special kind of RNN, which are designed for learning long-term dependencies. Recurrent neural networks have the form of a chain of repeating modules of neural networks,which can be as simple as a single tanh layer.
The Architecture of LSTMs is shown below in Figure 6. (Socher, 2015).

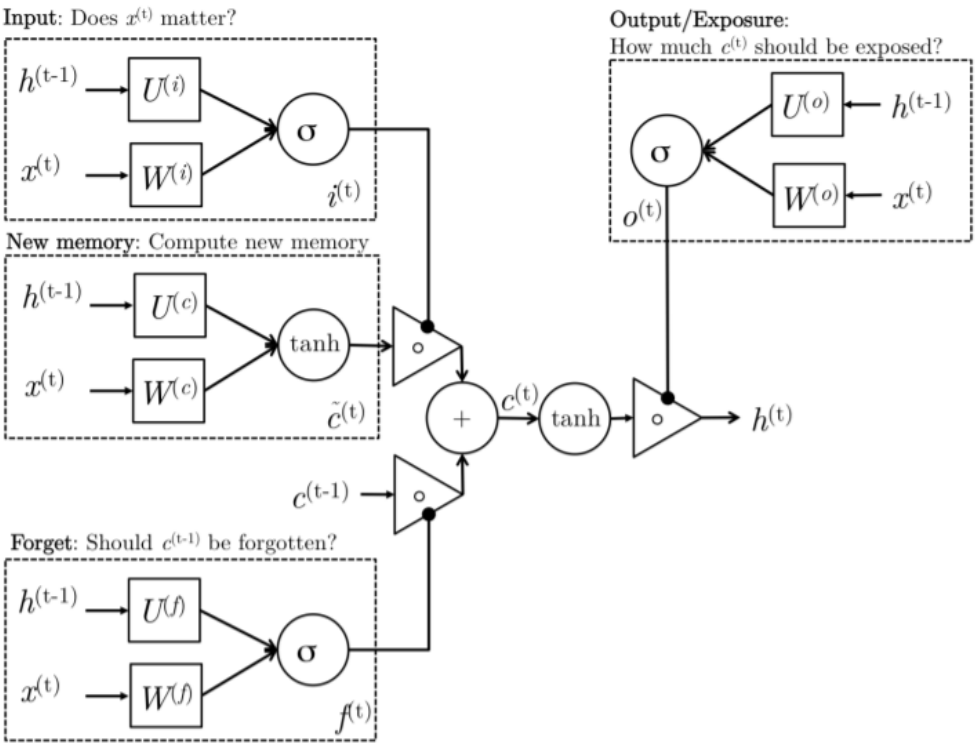


Figure 6 : Architecture of LSTMs

LSTMs can be better understood by considering their architecture in the following stages:

1. New memory generation: In this step, we use the input word $x^{(t)}$ and the past hidden state $h^{(t-1)}$ to generate a new memory $\tilde{c}^{(t)}$ which includes aspects of the new word $x^{(t)}$.

2. Input Gate: In the new memory generation stage, we did not check if the new word is even important before generating the new memory. The function of the input gate is to receive the

input word and use the past hidden state to determine whether the input is important or not. It thus produces $i^{(t)}$ as an indicator of this information.

3. Forget Gate: This gate is similar to the input gate except that it does not check for usefulness of the input word, instead it evaluates whether the past memory cell is useful for the computation of the present memory cell. Thus, the forget gate looks at the input word and the past hidden state and produces $f^{(t)}$.

4. Final memory generation: In this step, the model first takes the output of the forget gate $f^{(t)}$, and accordingly forgets the past memory $c^{(t-1)}$.Then it takes the output of the input gate $i^{(t)}$, and accordingly gates the memory $c'^{(t)}$. Then these two results are summed up to generate the final memory $c^{(t)}$.

5. Output/Exposure Gate: The final memory $c^{(t)}$ generated contains a large amount of data , some of which need not be saved in the hidden state. Hidden states are used in every gate of an LSTM, and hence it is important that only the relevant data is stored there. The Output Gate decides what aspects of the memory $c^{(t)}$ need to be saved in the hidden state $h^{(t)}$, and produces the signal $o^{(t)}$, which is used to gate the pointwise tanh of the memory.

e. Bi-directional Long Short Term Memory networks(Bi-LSTMs) - Bidirectional long-short term memory (Bi-LSTM) is any neural network capable of storing sequence information in both forward (past to future) and backward directions (future to past). It is similar to an LSTM model in most aspects, except that it has additional layers that enable the input to run in two directions, thus distinguishing it from a conventional LSTM.

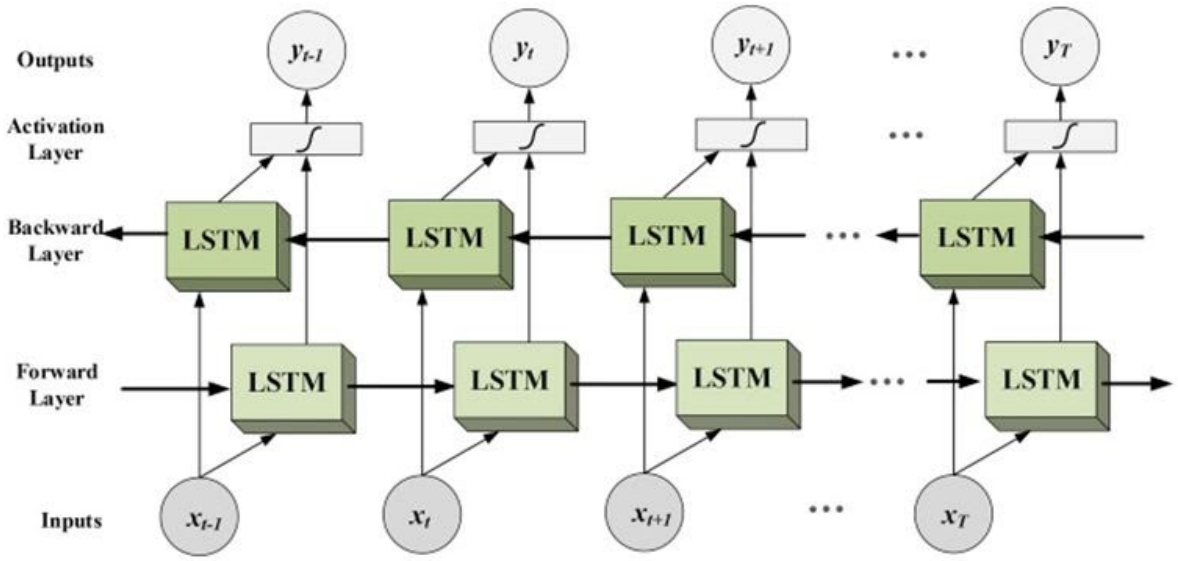


Figure 7: Architecture of a Simple Bi-LSTM Model (Verma, 2021)

Through the diagram we get an overall idea of the flow of information in backward and forward directions. Bi-LSTMs are typically used when sequence-to-sequence tasks are required, and since detecting Fake News falls within this category, Bi-LSTMs can be quite useful in this regard.

f. Ensemble Learning Methods -It has been observed that sometimes an algorithm can accurately classify fake news while most other algorithms fail to do the same. This is because that algorithm focuses on certain aspects of the text that others do not. So, if we aggregate these models together guided by some algorithms, we may arrive at a better model.

Two Major algorithms for this purpose are:

- Bagging: It is a homogeneous weak learners' model that learns from each other independently in parallel and combines them for determining the model average.
- Boosting: It is also a homogeneous weak learners' model but works differently from Bagging. In this model, learners learn sequentially and adaptively to improve model predictions of a learning algorithm.

A representation of their workflows have been shown in Figure 8(A Practical Tutorial on Bagging and Boosting Based Ensembles for Machine Learning: Algorithms, Software Tools, Performance Study, Practical Perspectives and Opportunities, 2020).

There is a meta-algorithm Stacking, which is a hybrid method that learns different models in parallel and combines them by training a meta-model to output predictions based on the different constituent models' predictions. Such a deterministic aggregation is expected to build a model that produces more tangible outputs and satisfactory results with much higher accuracy.

A Concept diagram of Stacking has been shown in Figure 9 (Singh, 2019).

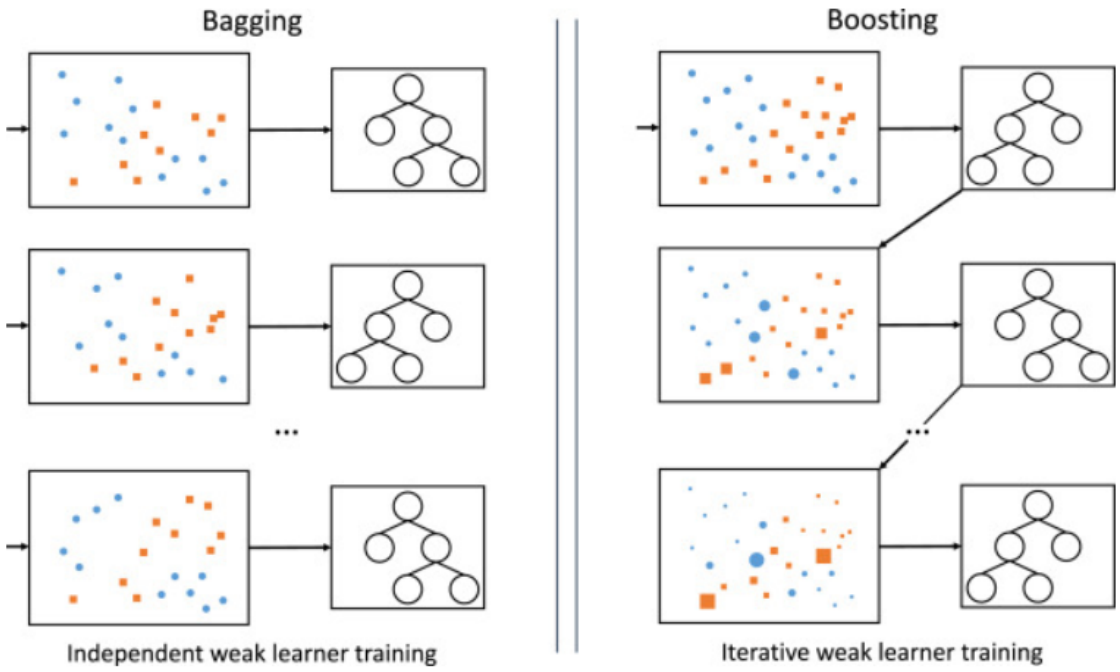


Figure 8: Representation of the workflows of Bagging and Boosting strategies.

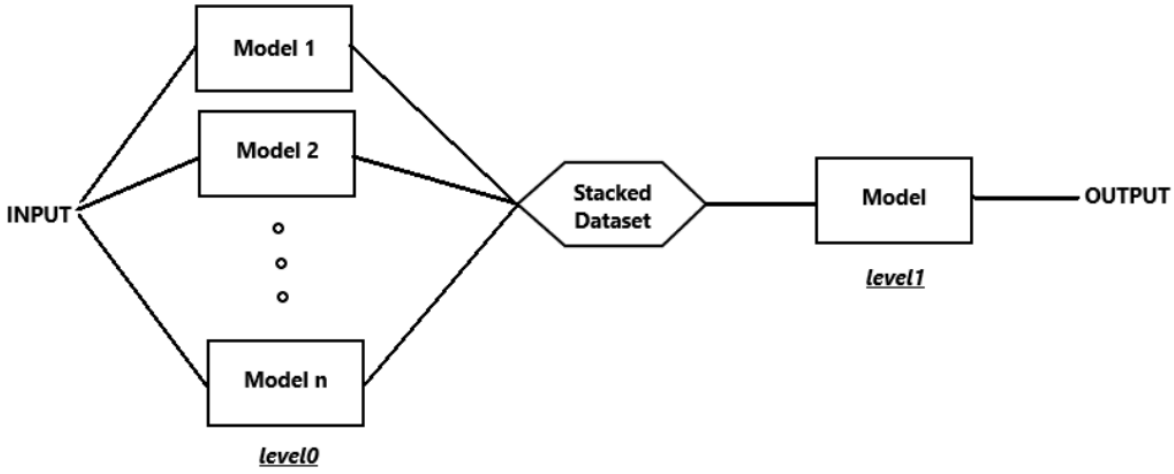


Figure 9: Concept Diagram of Stacking

Blending is another ensemble technique that can help us to improve performance and increase accuracy. It follows the same approach as stacking but uses only a holdout/ validation set from the train set to make predictions. Thus unlike stacking, the predictions are made on the holdout set itself, which are then used to build a model. (By Great Learning Team -, 2022)

An overview of the blending process is given below:

- The train set is divided into two sections: training and validation sets.
- The model(s) are fitted to the training set.
- Predictions are made on both the validation and test sets.
- The validation set and its predictions serve as features in the building of a new model.
- The final predictions on the test and meta-features are made using this model.

In the ensemble models we have tested, we have used Blending as our modes of implementation.

g. Random Forest Algorithm - Random forest algorithm is a machine learning algorithm based on ensemble learning that uses the process of combining various classifiers to enhance the performance of the model overall. It uses the results of numerous decision trees of various

subsets from the dataset given which is then averaged or voted upon to produce accurate results. The higher number of forest trees generally increases the Accuracy and Precision of the solution.

The most important asset of Random forests is that they maintain the level of accuracy even if the dataset has missing elements. It can be applied to a varied range of regression problems and prediction problems, since it takes fewer parameters to produce outputs and deals with complex datasets having higher dimensions as well. The algorithm is primarily used for classification and prediction of univariate and multivariate time series.
(Gaurkar S. et. al, 2021)

The steps to implement them are as follows:

- For each sample set, a decision tree is built.
- Every decision tree's prediction result is obtained and stored.
- Every prediction receives a vote.
- The most voted prediction result is selected as the final result.

A conceptual diagram has been shown in Figure 10.(Bakshi, 2021)

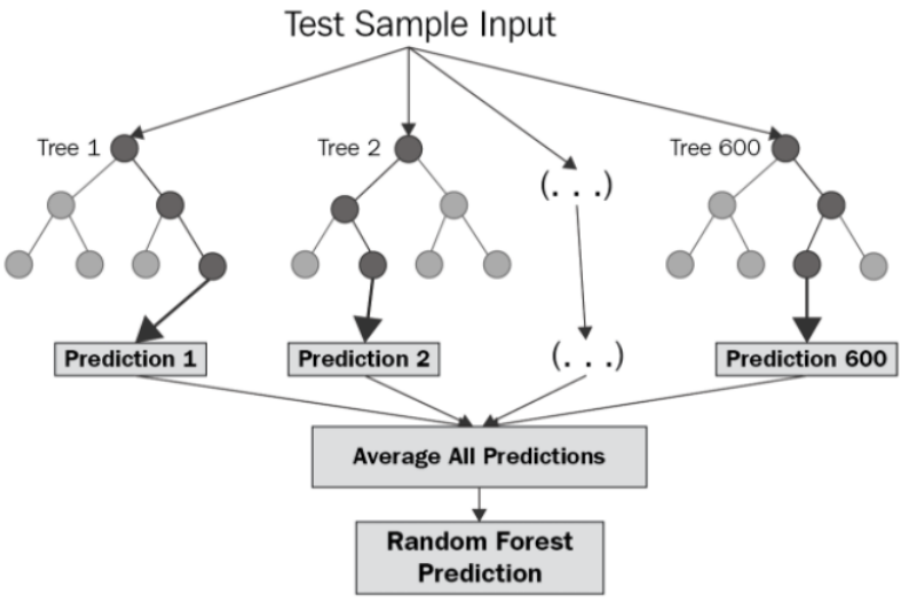


Figure 10: Concept Diagram of Random Forest Regression

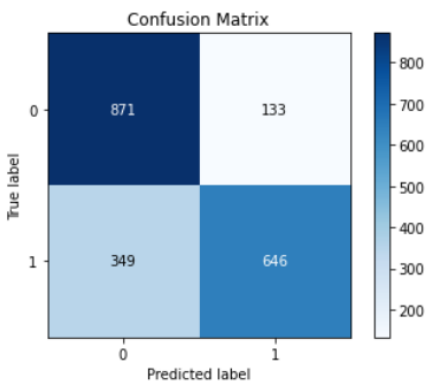
5.2.2. Analysis:

The dataset and the codes used for the models implemented are included in the appendix. The same dataset was used to evaluate all the models, and the results are tabulated below:

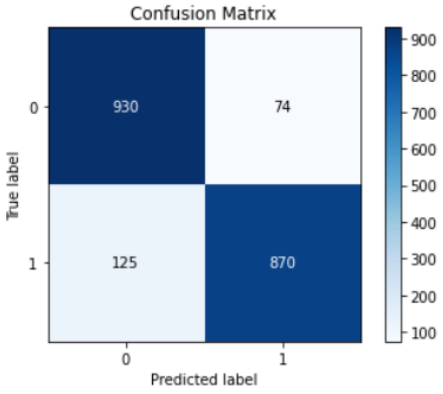
a. Comparing Performances of Models

Confusion Matrix

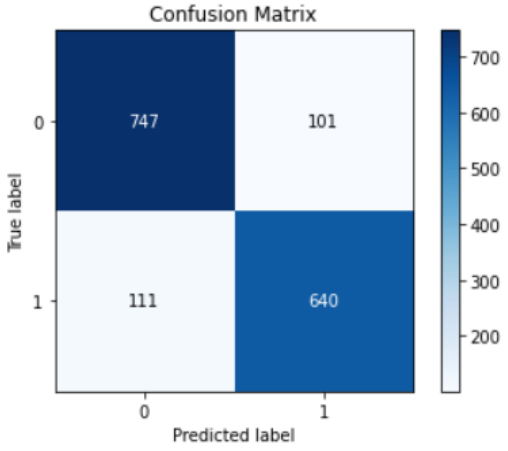
(Note: Here, 0 means Reliable News and 1 means Unreliable News)



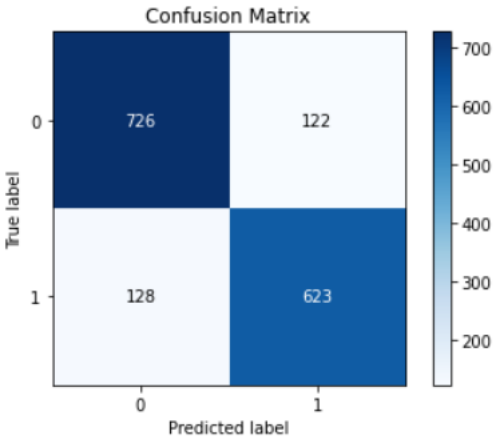
i. Naïve Bayes



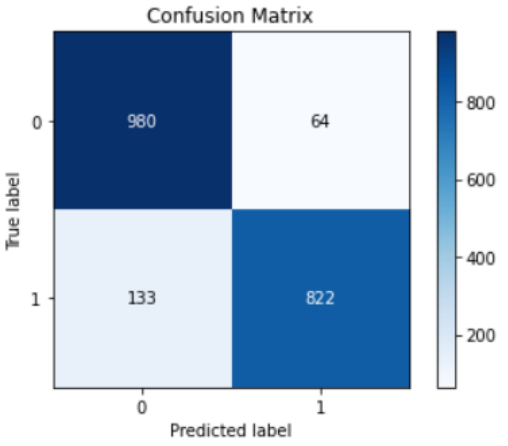
ii. SVM



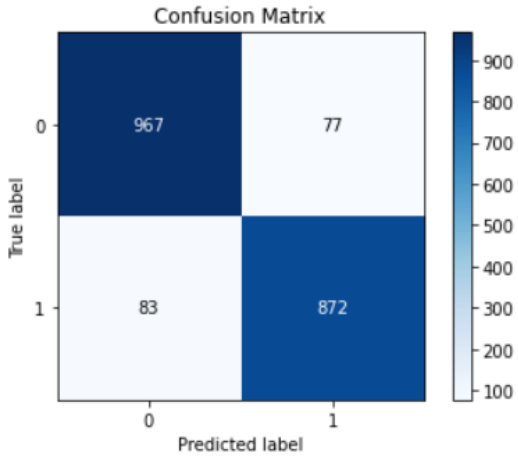
iii. NN with Keras



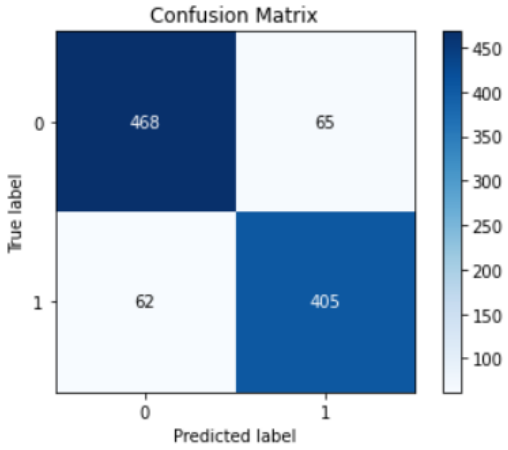
iv. Random Forest Regression



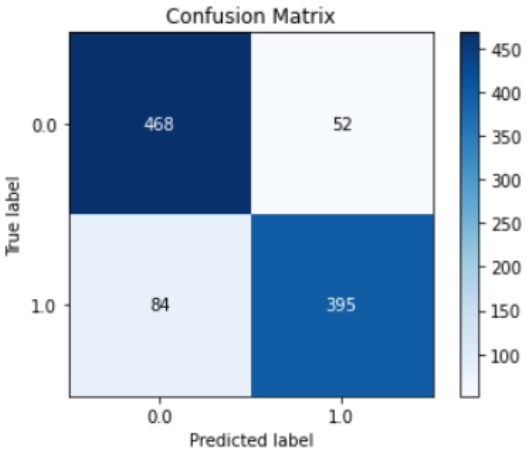
v. LSTM



vi. Bi-LSTM



vii. Memory-Based Ensemble Model



viii. Non Memory-Based Ensemble Model

Table 2: Confusion matrices of different models

We can make the following deductions based on the confusion matrix for

a) Naives Bayes:

(i) 646 fake news articles have been correctly predicted as fake

(ii) 871 news articles that are true (real) have been correctly predicted as true

(iii) 349 fake news articles have been incorrectly predicted as true

(iv) 133 true news articles have been incorrectly predicted as fake

b) Support Vector Machine:

(i) 870 fake news articles have been correctly predicted as fake

(ii) 930 news articles that are true (real) have been correctly predicted as true

(iii) 125 fake news articles have been incorrectly predicted as true

(iv) 74 true news articles have been incorrectly predicted as fake

c) Neural Network with Keras:

(i) 640 fake news articles have been correctly predicted as fake

(ii) 747 news articles that are true (real) have been correctly predicted as true

(iii) 111 fake news articles have been incorrectly predicted as true

(iv) 101 true news articles have been incorrectly predicted as fake

d) Random Forest Regression:

(i) 623 fake news articles have been correctly predicted as fake

(ii) 726 news articles that are true (real) have been correctly predicted as true

(iii) 128 fake news articles have been incorrectly predicted as true

(iv) 122 true news articles have been incorrectly predicted as fake

e) Long Short Term Memory:

(i) 822 fake news articles have been correctly predicted as fake

(ii) 980 news articles that are true (real) have been correctly predicted as true

(iii) 133 fake news articles have been incorrectly predicted as true

(iv) 64 true news articles have been incorrectly predicted as fake

f) Bi-directional Long Short Term Memory:

(i) 872 fake news articles have been correctly predicted as fake

(ii) 967 news articles that are true (real) have been correctly predicted as true

(iii) 83 fake news articles have been incorrectly predicted as true

(iv) 77 true news articles have been incorrectly predicted as fake

g) Memory-Based Ensemble Model:

(i) 405 fake news articles have been correctly predicted as fake

(ii) 468 news articles that are true (real) have been correctly predicted as true

(iii) 62 fake news articles have been incorrectly predicted as true

(iv) 65 true news articles have been incorrectly predicted as fake

h) Non Memory-Based Ensemble Model:

(i) 395 fake news articles have been correctly predicted as fake

(ii) 468 news articles that are true (real) have been correctly predicted as true

(iii) 84 fake news articles have been incorrectly predicted as true

(iv) 52 true news articles have been incorrectly predicted as fake

Parameters→/ Models↓	Accuracy	F1-Score	Recall	Precision	Specificity
Naïve Bayes	*75.89%*	*72.83%*	*64.92%*	*82.93%*	86.75%

Random Forest Regression	84.37%	83.29%	82.96%	83.62%	*85.61%*
SVM	90.05%	89.74%	87.44%	92.16%	92.63%
NN with Keras	86.74%	85.79%	85.22%	86.37%	88.09%
LSTM	90.15%	89.30%	86.07%	**92.78%**	**93.87%**
Bi-LSTM	**92.00%**	**91.60%**	**91.31%**	91.89%	92.62%
Non Memory-Based Ensemble Model	86.39%	85.31%	82.46%	88.37%	90.00%
Memory-Based Ensemble Model	87.30%	86.45%	86.72%	86.17%	87.80%

Table 3: Comparison among performances of different models

Thus, we have the following observations from the outputs produced by the models:

1. Naive Bayes is constantly the least effective of all in all aspects, showing that we cannot consider all words of equal importance, and weights must be assigned to words.
2. SVMs performed quite well , almost as well as the Memory based models. It even outperformed Bi-LSTMs in terms of Precision and Specificity, however was outperformed by LSTMs in all aspects.
3. Random Forest Regressors, which act as ensemble models in themselves, performed almost as well as NN with Keras, but were observed to be less efficient in comparison.

4. LSTMs and Bi-LSTMs are the most effective in detecting fake news since they add the aspect of memory into neural networks using recurrent neural networks. Bi-LSTM did better than LSTM in most aspects, falling short in terms of Precision and Specificity.
5. The introduction of Non-Memory based ensemble models did nothing significant to improve the performance of the models and still remained lower than the memory based models LSTM and Bi-LSTM in all aspects.
6. A similar trend was observed in the case of Memory based ensemble model as there too no significant improvement was observed. However, the Non-Memory based ensemble model did manage to outdo the Memory based ensemble model in terms of Precision and Specificity.

These observations have been obtained after we ran the models multiple times and also have been found to be in line with the results obtained in other similar studies. Although it was seen that SVMs performed quite well, the observations show that the more intensive we get into neural networks and introduce more aspects like memory, the more accurate the models become. Another level of introduction of Ensemble models was expected to further improve the performances of both memory and non memory based models, but in our case, it was not so prominent and the models performed better off alone.

Overall, it can be seen that AI has helped a lot in providing hints regarding the binary classification and thus is quite effective in classifying fake news. This stands in support of our Research Hypothesis that Memory based models do in fact improve the accuracy of detecting fake news propagated on social media. However in our study, Ensemble Models did not show any significant improvements in both the cases of Memory based as well as Non-Memory based models.

b. Current Methods - Statista is a German company specializing in market and consumer data[2]. As of October 2021, according to Statista (Statista, 2021), Facebook has the largest number of

[2]https://en.wikipedia.org/wiki/Statista#:~:text=Statista%20is%20a%20German%20company,of%20about%20%E2%82%AC60%20million.

active users online, which naturally means it is the hub of production as well as propagation of fake news and misinformation pertaining to all sorts of topics.Facebook maintains an official blog where they brief the various AI methods they use on their platforms. As per the blog, although Facebook has employed multiple policies and products to solve various challenges and contain misinformation as much as possible, the increasing spread of fake news has brought to light a major technical challenge: Image Manipulation. A growing concern today is Deepfakes. Deepfakes are synthetic media in which a person in an existing image or video is replaced with someone else's likeness. However, this problem is not limited to people; an image of just about anything being shared by many online may have been manipulated in a way to give a message radically different from what was intended. (Facebook AI, 2021)[3].

SimSearchNet is a convolutional neural network model built based on a multiyear collaboration by Facebook AI researchers, engineers, and many others across the company. Currently, Facebook has deployed SimSearchNet++, an enhanced image matching model trained using unsupervised learning to track variations of an image with a high degree of precision and improved recall. It runs on images uploaded to Facebook and Instagram, and is resistant to crops, blurs, and screenshots, among other image manipulations. For images containing text, SimSearchNet++ can also group matches with high precision using optical character recognition (OCR) verification, ensuring no aspect of images go unchecked at all.

Another challenge in detecting misinformation is that different fake news articles may contain the same information and motive, but express them in very different ways by means of rephrasing the articles, using different images, or changing the format from graphic to text. Facebook is currently implementing new AI systems that automatically detect new variants of content already discredited by independent fact-checkers. When the AI model detects such new variants, they are flagged and forwarded to their fact-checking partners for review.

These advents have enabled Facebook to predict more matches and identify fake news faster, significantly inhibiting the spread of misinformation on the Social Media giant.

c. Future Directions of Study -

A disadvantage of our study is that since our current models are majorly based on Supervised Learning, they require large amounts of data sets from reliable sources- both for

[3] https://ai.facebook.com/blog/heres-how-were-using-ai-to-help-detect-misinformation

training and testing models, and these datasets should not be restricted to a particular domain, like politics. Such datasets are not readily available on the Internet and also cost a significant amount of time to compile and verify. So a switch to an Unsupervised or at least a Semi-supervised learning approach would be beneficial. The basic idea is to access various social media platforms, follow multiple users who have posted regarding a particle article, and arrive at a conclusion based on the data collected from the users there. However, simply following what is dictated by the majority of votes in a poll(referred to as Crowdsourcing) will not always be the best idea since most people may be under a wrong impression and may also spread the same fake news widely.

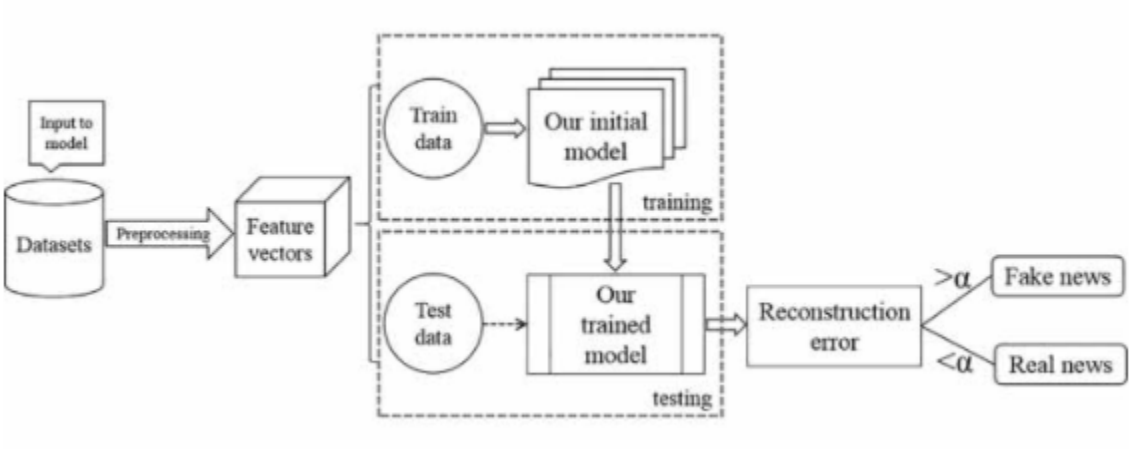


Figure 9: Unsupervised Learning and Neural Network for News Classification

To overcome this problem, we can apply certain conditions to the users analyzed. The presence of a verified account on the social media platform, history of suspensions and bans, frequency of posting, the likes and dislikes on their content, comments on their posts, and the amount of time for which the user has been active on the platform are among the conditions that can be imposed on the users to check their credibility. After assigning weights based on these features, we can fuse these user features with the text features as is being done by previous models and once again run the model on test datasets, and the results obtained are predicted to be of much higher accuracy than those provided by models in use now (*Unsupervised Fake News Detection Based on Autoencoder*, 2021). An outline for the implementation of this approach has been shown in Figure 9. SimSearchNet++ used by Facebook as an enhanced image matching model is based on this method(Facebook AI, 2021).

Also, since the model does not follow a user-provided dataset and relies on real-time analysis, the model should not require constant upgrades. Instead, it would be able to train itself to accurately detect fake news, even as all the features of the news text - including the style, language, motive, and means of fake news, change with time.

It has also been observed that models generally perform worse when analyzing news in local languages. However, in some cases, credibility ratings obtained from crowdsourcing have proven to be quite effective. Hence for local scenarios, the training of the models can be aided by updating the data with labels from crowdsourcing to improve their accuracies. (Pennycook G, Rang DG, 2019).

5.3. Conclusion

For our first objective, we circulated a questionnaire to collect information about the existing news habits of two age groups- one targeting University students in the 18-22 years age group (Students) and another above 30 years of age (Adults) via WhatsApp groups. Based on the results of our questionnaire, we found that irrespective of the age group, social media is a significant source of information for everyone. YouTube and WhatsApp were seen to be the most commonly used by both the Demographic groups. While both groups stated that they came across fake news, neither reported actually verifying the news they obtained. So apart from inculcating a news verification habit in the society, to curb fake news and misinformation, we need an efficient method integrated within these social media platforms to control propagation of fake news on social media. Flagging fake news has been successful in controlling the spread of misinformation greatly and due to the large scale of this problem, automated solutions like ML based classifiers are necessary to accomplish this task. This calls for a need to develop and test various AI models to come up with one that is most effective.

Thus for our second objective, we focussed mainly on the methodology for creating a fake news detector that can predict the fake news articles accurately. Our methodology particularly analyzed the article content and tried identifying key features in the text to achieve an effective classification of fake news. The prediction is implemented with the use of 5 models we made using Naive Bayes algorithm, SVMs, Neural Networks, LSTM and Bi-LSTM , and

then ensemble models based on Regression Forest Algorithms, the memory based and non-memory based models to produce accurate predictions of the fake news using the input variables.

From the results obtained in the previous section, we observed that Naive Bayes performed the worst and while SVMs too performed quite well but Memory-based models (LSTMS and Bi-LSTMs) overall worked best in classifying fake news. In our study, Ensemble Models did not show any significant improvements in either case. The study overall supported our Research Hypothesis that Memory based models do in fact improve the accuracy of detecting fake news propagated on social media.

Regarding current methods used, SimSearchNet- an initiative by Facebook, has been successful to a great extent in not only detecting text based fake news, but also those involving graphics. OCR Technology can be used to extract any text that may be present and along with image classification, more accurately identify fake news.

Facebook also has other algorithms to check if the same piece of news has been represented in some other way trying to mislead people and flag those for review from their fact-checking partners as well. The exact methods used by Facebook and other Social Media platforms are not available officially, so it is difficult to comment what are the exact methods used.

But it is clear that the Future areas of study include the incorporation of Multi-Model Approaches, and a transition to Unsupervised Training Models so that even without the modification of used algorithms, the models themselves can evolve and adapt themselves to changes so as to be able to adapt to changes in features of fake news and successfully stand the test of time.

5.4. Limitations

Because we didn't have any outside funding, we had to identify variables and dimensions

based on our literature review, and survey which too was limited due to the times of COVID-19. Hence, we were unable to collect our own data and had to rely more on secondary sources.

The dataset used to train our models was secondary and so the dataset available was limited in quantity. Moreover, the news articles were random in nature rather than focussed on a political or social context. So, a more specific, larger and detailed primary dataset is needed to train models better in the future.

6. ACKNOWLEDGEMENTS

We would like to thank Dr. Tanu Shukla, Department of Humanities and Social Sciences, BITS Pilani, for her guidance, support, encouragement, and valuable critiques in this research project. We would also like to thank Ms. V Mounika Prashanthi and Ms. Sugandha Bhatnagar for constantly guiding and helping us throughout our project.

7. REFERENCES

A practical tutorial on bagging and boosting based ensembles for machine learning: Algorithms, software tools, performance study, practical perspectives and opportunities. (2020, December 1). ScienceDirect. https://www.sciencedirect.com/science/article/pii/S1566253520303195

Albahar, M. (2021). A hybrid model for fake news detection: Leveraging news content and user comments in fake news. IET Information Security, 15(2), 169–177. https://doi.org/10.1049/ise2.12021

Aldwairi, M., & Alwahedi, A. (2018). Detecting Fake News in Social Media Networks. *Procedia Computer Science*, *141*, 215–222. https://doi.org/10.1016/j.procs.2018.10.171

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. Journal of Economic Perspectives, 31(2), 211–236.

Ahmad, I., Yousaf, M., Yousaf, S., & Ahmad, M. O. (2020). Fake News Detection Using Machine Learning Ensemble Methods. *Complexity*, *2020*, 1–11. https://doi.org/10.1155/2020/8885861

Alonso García, S., Gómez García, G., Sanz Prieto, M., Moreno Guerrero, A. J., & Rodríguez Jiménez, C. (2020). The Impact of Term Fake News on the Scientific Community. Scientific Performance and Mapping in Web of Science. *Social Sciences*, *9*(5), 73. https://doi.org/10.3390/socsci9050073

Al-Zaman, M. S. (2021). Prevalence and source analysis of COVID-19 misinformation in 138 countries. *IFLA Journal*, 034003522110411. https://doi.org/10.1177/03400352211041135)

Apuke, O. D., & Omar, B. (2021). Fake news and COVID-19: modelling the predictors of fake news sharing among social media users. *Telematics and Informatics*, *56*, 101475. https://doi.org/10.1016/j.tele.2020.101475

Bahad, P., Saxena, P., & Kamal, R. (2019). Fake News Detection using Bi-directional LSTM-Recurrent Neural Network. Procedia Computer Science, 165, 74–82. https://doi.org/10.1016/j.procs.2020.01.072

Bakshi, C. (2021, December 14). Random Forest Regression - Level Up Coding. Medium. https://levelup.gitconnected.com/random-forest-regression-209c0f354c84

BALI, A., & DESAI, P. (2019). Fake News and Social Media: Indian Perspective. *Media Watch*, *10*(3). https://doi.org/10.15655/mw/2019/v10i3/49687

Barthel, M., Mitchell, A., & Holcomb, J. (2016). Many Americans believe fake news is sowing confusion. Pew Research Center. Retrieved from http://www.journalism.org/2016/12/15/manyamericans-believe-fake-news-is-sowingconfusion/

By Great Learning Team -. (2022, March 22). Ensemble learning with Stacking and Blending | What is Ensemble Learning? GreatLearning Blog: Free Resources What Matters to Shape Your Career! https://www.mygreatlearning.com/blog/ensemble-learning/

Dasaradh, S. K. (2020). *A Gentle Introduction To Math Behind Neural Networks*. Towardsdatascience. https://towardsdatascience.com/introduction-to-math-behind-neural-networks-e8b60dbbdeba

Duffy, A., Tandoc, E., & Ling, R. (2019). Too good to be true, too good not to share: the social utility of fake news. *Information, Communication & Society*, *23*(13), 1965–1979. https://doi.org/10.1080/1369118x.2019.1623904

Egelhofer, J. L., & Lecheler, S. (2019). Fake news as a two-dimensional phenomenon: a framework and research agenda. *Annals of the International Communication Association*, *43*(2), 97–116. https://doi.org/10.1080/23808985.2019.1602782

Flostrand, A., Pitt, L., & Kietzmann, J. (2019). Fake news and brand management: a Delphi study of impact, vulnerability and mitigation. *Journal of Product & Brand Management*, *29*(2), 246–254. https://doi.org/10.1108/jpbm-12-2018-2156

Facebook AI, 2021. Here's how we're using AI to help detect misinformation. Facebook AI. (n.d.). Retrieved December 2, 2021, from https://ai.facebook.com/blog/heres-how-were-using-ai-to-help-detect-misinformation.

Gaozhao, D. (2020). Flagging Fake News on Social Media: An Experimental Study of Media Consumers' Identification of Fake News. *SSRN Electronic Journal*. Published. https://doi.org/10.2139/ssrn.3669375

Giuliani-Hoffman, Francesca (November 3, 2017). "'F*** News' should be replaced by these words, Claire Wardle says". Money.CNN. Retrieved November 24, 2018.

H. Rashkin, E. Choi, J. Y. Jang, S. Volkova, and Y. Choi, "Truth of varying shades: Analyzing language in fake news and political fact-checking," EMNLP 2017 - Conf. Empir. Methods Nat. Lang. Process. Proc., pp. 2931–2937, 2017, doi: 10.18653/v1/d17- 1317.

Hansrajh, A. (2021, July 29). *Detection of Online Fake News Using Blending Ensemble Learning*. https://www.hindawi.com/journals/sp/2021/3434458/

Jonathan Albright, J. A. (2017, June 27). *Welcome to the Era of Fake News*. www.Cogitatiopress.com. Media and Communication. Published. https://www.cogitatiopress.com/mediaandcommunication/article/view/977/977

Lee, L. W., Hannah, D., & McCarthy, I. P. (2019). Do your employees think your slogan is “fake news?” A framework for understanding the impact of fake company slogans on employees. *Journal of Product & Brand Management*, *29*(2), 199–208. https://doi.org/10.1108/jpbm-12-2018-2147

Liu, Y., & Wu, Y. F. B. (2020). FNED. ACM Transactions on Information Systems, 38(3), 1–33. https://doi.org/10.1145/3386253

Mahabub, A. (2020). A robust technique of fake news detection using Ensemble Voting Classifier and comparison with other classifiers. SN Applied Sciences, 2(4). https://doi.org/10.1007/s42452-020-2326-y

Marr, B.: Coronavirus fake news: how Facebook, Twitter, and Instagram are tackling the problem. Forbes (2020).

Mengji, S. (2021). Fake News Detection using RNN-LSTM. *International Journal for Research in Applied Science and Engineering Technology*, *9*(10), 1731–1737. https://doi.org/10.22214/ijraset.2021.35687

Nagi, Kuldeep, New Social Media and Impact of Fake News on Society (June 6, 2018). ICSSM Proceedings, July 2018, Chaing Mai, Thailand, pp. 77-96, Available at SSRN: https://ssrn.com/abstract=3258350

Pennycook G, Rand DG. Fighting misinformation on social media using crowdsourced judgments of news source quality. *Proc. Natl. Acad. Sci.* 2019;116(7):2521–2526. doi: 10.1073/pnas.1806781116. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

Rocha, Y. M., de Moura, G. A., Desidério, G. A., de Oliveira, C. H., Lourenço, F. D., & de Figueiredo Nicolete, L. D. (2021). The impact of fake news on social media and its influence on health during the COVID-19 pandemic: a systematic review. *Journal of Public Health*. Published. https://doi.org/10.1007/s10389-021-01658-z

S. A. Khan, M. H. Alkawaz and H. M. Zangana, "The Use and Abuse of Social Media for Spreading Fake News," 2019 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS), 2019, pp. 145-148, doi: 10.1109/I2CACIS.2019.8825029.

S. Aphiwongsophon and P. Chongstitvatana, "Detecting Fake News with Machine Learning Method," 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, 2018.

S. Gaurkar , A. Kotalwar, S. Gabale (November, 2021). “Predictive Maintenance of Industrial Machines using Machine Learning” . *International Research Journal of Engineering and Technology (IRJET)*. Published. https://www.irjet.net/archives/V8/i11/IRJET-V8I1186.pdf

Sastrawan, I. K., Bayupati, I., & Arsa, D. M. S. (2021). Detection of fake news using deep learning CNN–RNN based methods. ICT Express. Published. https://doi.org/10.1016/j.icte.2021.10.003

Singh, S. P. (2019, July 29). Understand Stacked Generalization (blending) in depth with code demonstration. OpenGenus IQ: Computing Expertise & Legacy. https://iq.opengenus.org/stacked-generalization-blending/

Socher, R. (2015). *Https://cs224d.stanford.edu/lecture_notes/LectureNotes4.pdf*. *www.stanford.edu*. Retrieved from https://cs224d.stanford.edu/lecture_notes/LectureNotes4.pdf.

Statista. (2021, November 16). *Global social networks ranked by number of users 2021*. https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/

Tandoc, E. C., Lim, D., & Ling, R. (2019). Diffusion of disinformation: How social media users respond to fake news and why. *Journalism*, *21*(3), 381–398. https://doi.org/10.1177/1464884919868325

Unsupervised Fake News Detection Based on Autoencoder. (2021). IEEE Journals & Magazine | IEEE Xplore. https://ieeexplore.ieee.org/document/9352726

V. Pérez-Rosas, B. Kleinberg, A. Lefevre, and R. Mihalcea, "Automatic Detection of Fake News," 2017, [Online]. Available: http://arxiv.org/abs/1708.07104.

Verma, Y. (2021, November 20). Complete Guide To Bidirectional LSTM. Analytics India Magazine. https://analyticsindiamag.com/complete-guide-to-bidirectional-lstm-with-python-codes/

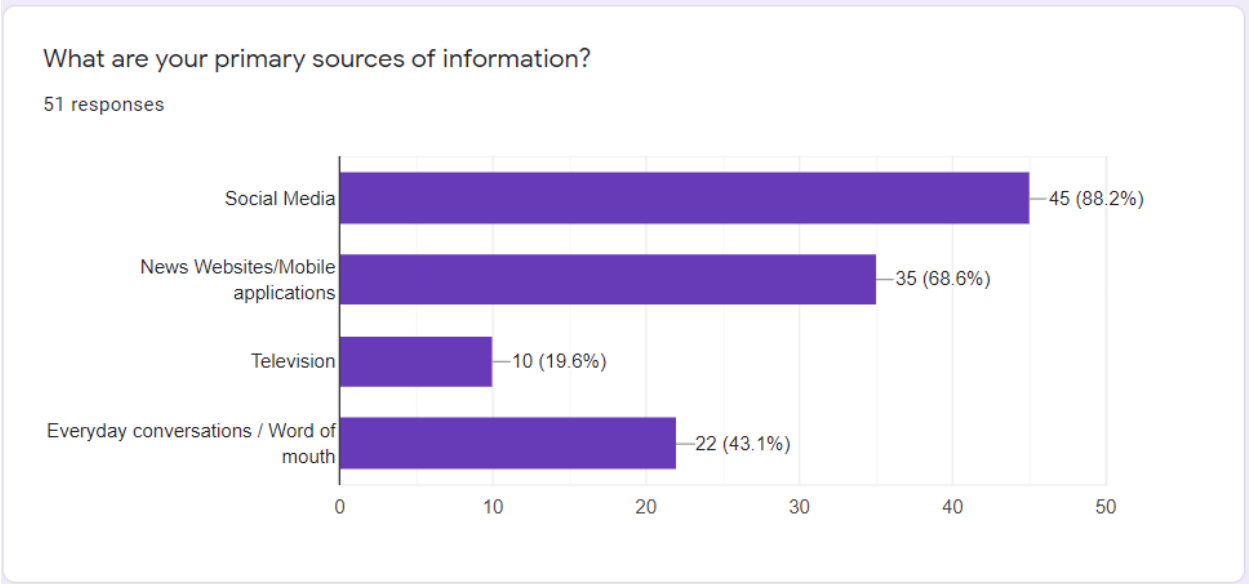
Verma, P. K., Agrawal, P., Amorim, I., & Prodan, R. (2021). WELFake: Word Embedding Over Linguistic Features for Fake News Detection. IEEE Transactions on Computational Social Systems, 8(4), 881–893. https://doi.org/10.1109/tcss.2021.3068519

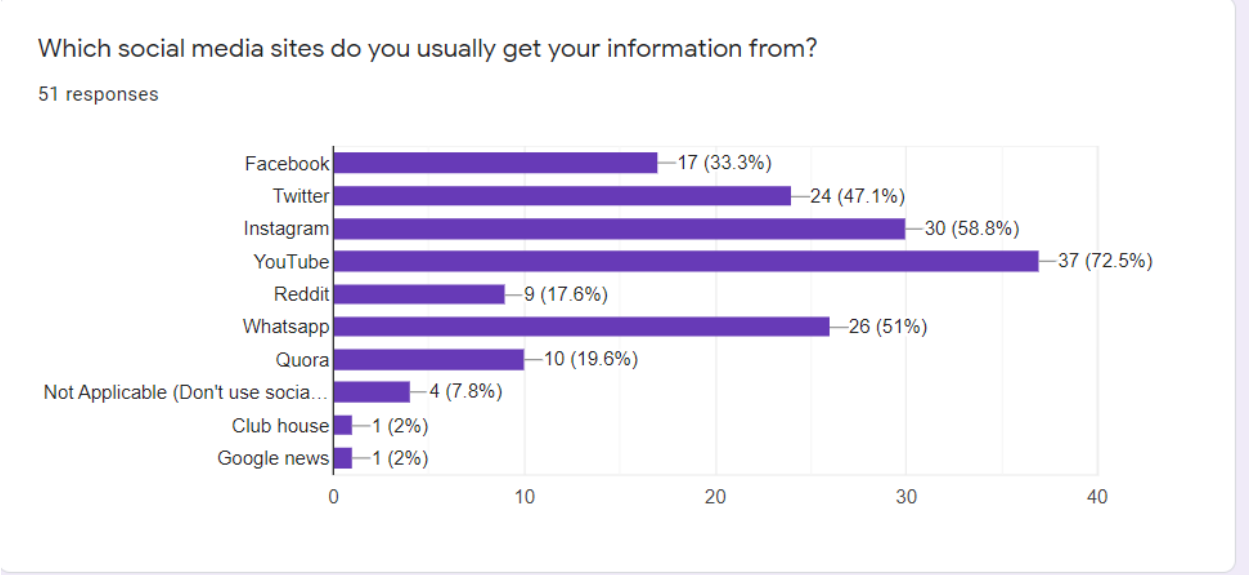
Vijayaraghavan S., Wang Y., Guo Z., Voong J., Xu W., Nasseri A., Cai J., Li L., Vuong K., & Wadhwa E. (2020) .. Cs.CL. Published.

Vij, S. (2018) *https://theprint.in/opinion/a-single-whatsapp-rumour-has-killed-29-people-in-india-and-nobody-cares/77634/* . Print. https://theprint.in

8. APPENDIX

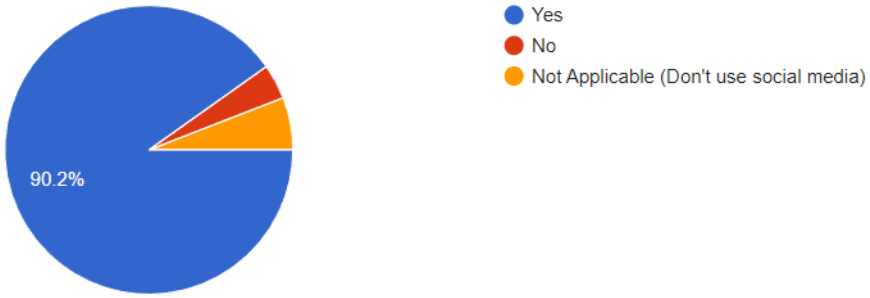
8.1. Results of Questionnaire For University Students (Age Group 18 to 22)-





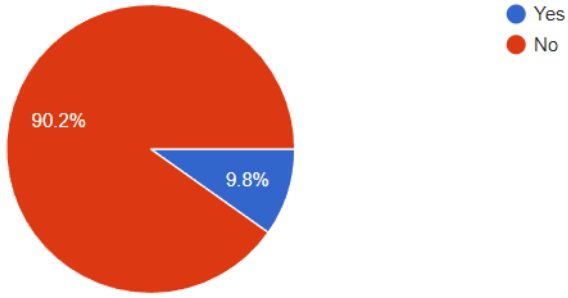
Have you seen these or similar labels while using social media?

51 responses



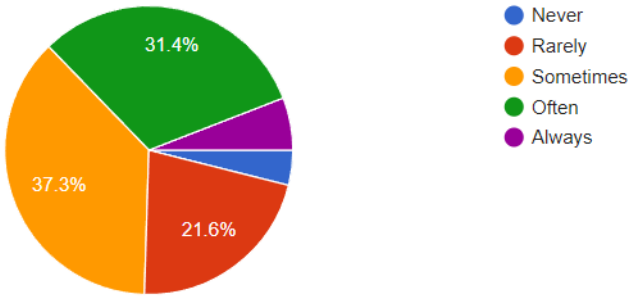
Do you pay for any digital subscriptions of news sources?

51 responses



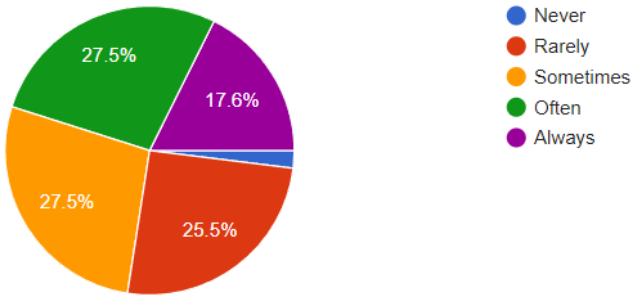
How frequently do you come across fake news?

51 responses



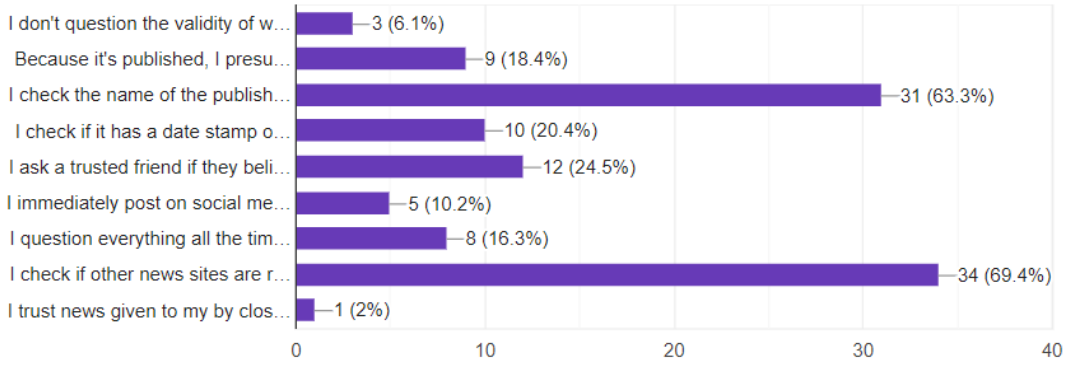
How frequently do you try verify the news you get from sources other than official channels?

51 responses



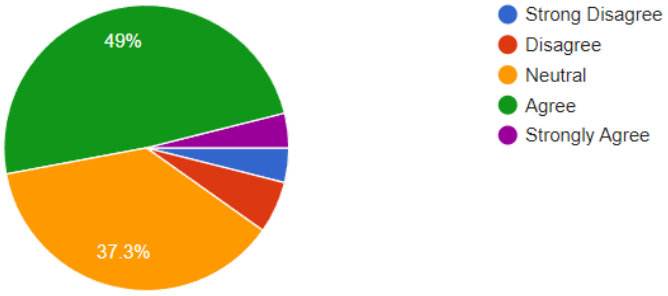
If you do tend to verify news/information, how/when do you do so?

49 responses



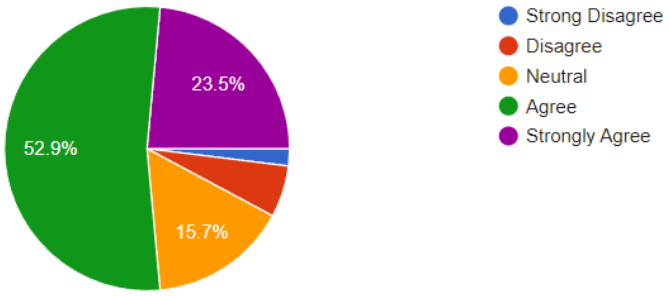
Technology and social media have made me a smarter and more informed person.

51 responses



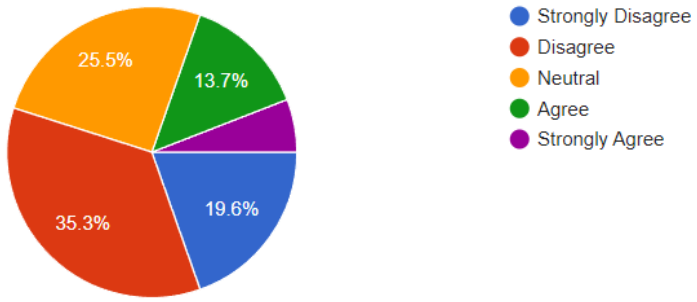
The lack of control and fact-checking on social media makes it suitable for the propagation of unconfirmed and/or incorrect information.

51 responses



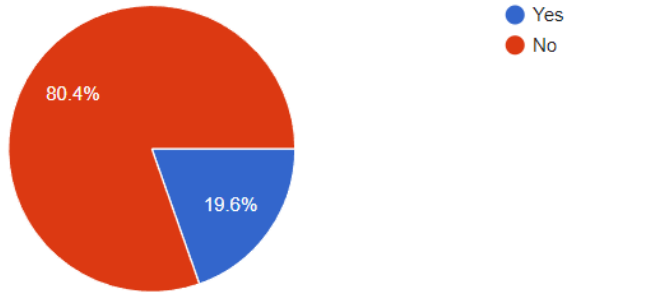
Traditional news outlets don't report fake news.

51 responses



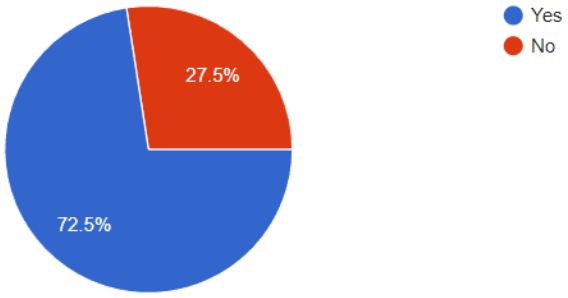
Do you use fact-checking websites like Snopes.com, FactCheck.org, PolitiFact.com and others?

51 responses



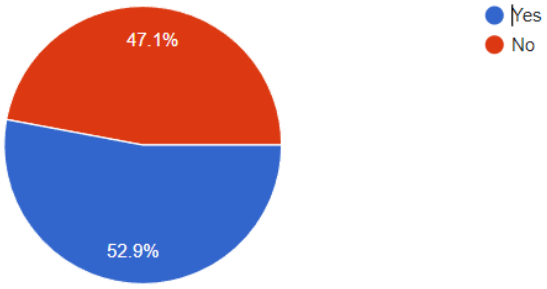
Did your colleagues ever inform you about an article being fake news?

51 responses



Have you ever been affected or seen someone affected due to circulation of fake news in any way?

51 responses



If your answer to the previous question is yes, then how?

8 responses

The case a few years back where two young men were mob lynched to death in Karbi Anglong, Assam, due to the spread of the fake news that they were kidnappers.

If you remember Freedom 251 phone news, one of biggest scams in India, it affected millions of people making them fall into the fraud.

Parents believing in fake news while i have to convince them it's fake

Spreading those fake Amazon gift ads

Affected reputation

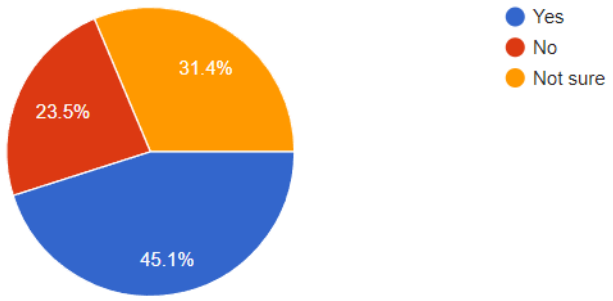
Reputation went down, people started avoiding me, making fun of me

I have seen many incidents where fake news has led to intercommunal fights

A girl in my highschool commited suicide bacause of the rumours people were spreading about her.

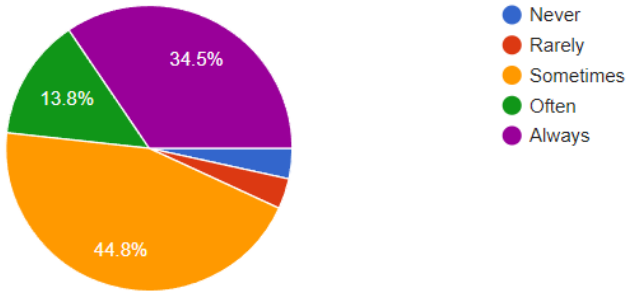
Has any of the news you have shared turned out to be fake ?

51 responses



If your answer is yes for the above question, do you share the corrected version afterwards?

29 responses



Do you have any suggestions regarding confirming authenticity of news?

8 responses

The simplest way would be a direct Google search that would give access of a wide array of information, from which it would be easier for us to discern the truth withheld.

And obvious grammatical errors clubbed with extremely hopeful/hopeless cases would definitely raise doubts, so in such cases it would be best to verify them instead of a self imposed blind faith on media.

As per me if the same news is being circulated by all outlets then it must be authentic. Of course everybody cannot spread fake news at the same time. So the best way can be by confirming the sme facts from multiple trusted sources.🙂🙂🙂🙂🙂🙂

Many popular traditional media is a sell out and runs according to propoganda of a particular political party, avoid those and read news from established newspaper media outlet (or media outlets which are not dependent on revenue generated by advertisements).

Check Wikipedia.

Not as such , but many a times news too good to be true are false

Do you have any suggestions regarding confirming authenticity of news?

8 responses

As per me if the same news is being circulated by all outlets then it must be authentic. Of course everybody cannot spread fake news at the same time. So the best way can be by confirming the sme facts from multiple trusted sources.🙂🙂🙂🙂🙂🙂

Many popular traditional media is a sell out and runs according to propoganda of a particular political party, avoid those and read news from established newspaper media outlet (or media outlets which are not dependent on revenue generated by advertisements).

Check Wikipedia.

Not as such , but many a times news too good to be true are false

Searching it on google and finding similar results

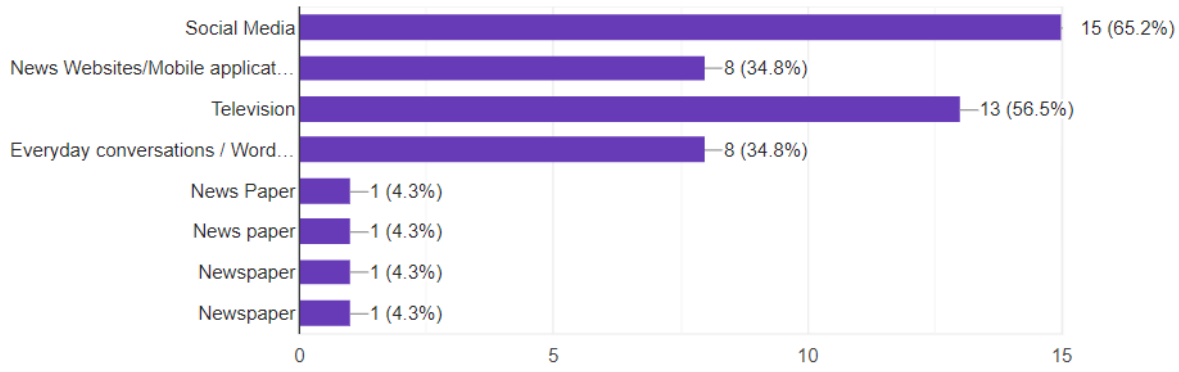
Check news in the news papers and magazines

Checking if other sites have the same news

8.2. Results Of Questionnaire For Adults (Age group 30+) -

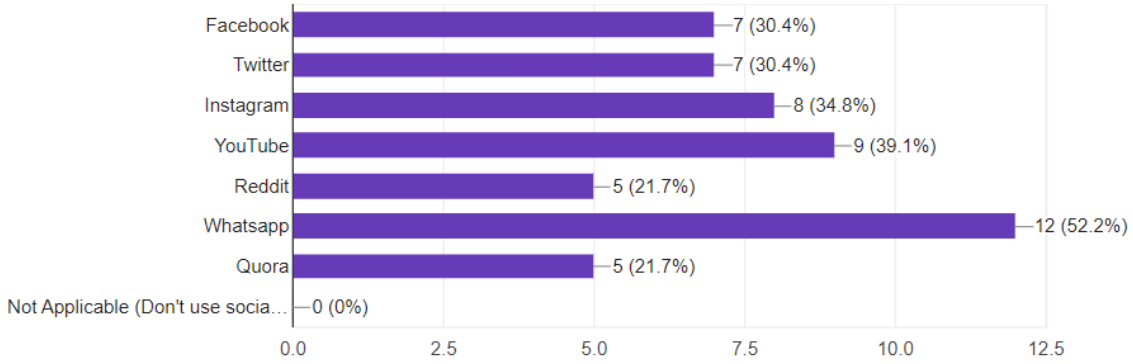
What are your primary sources of information?

23 responses



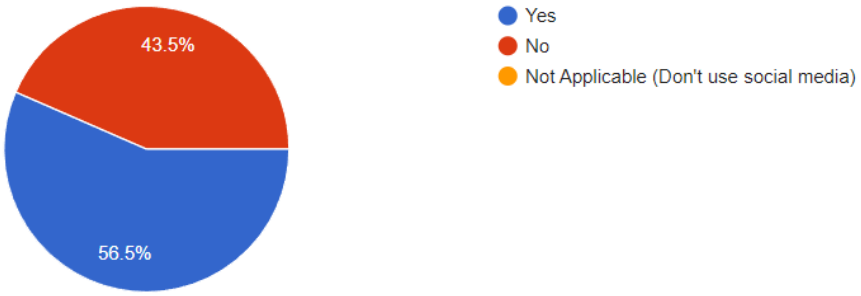
Which social media sites do you usually get your information from?

23 responses



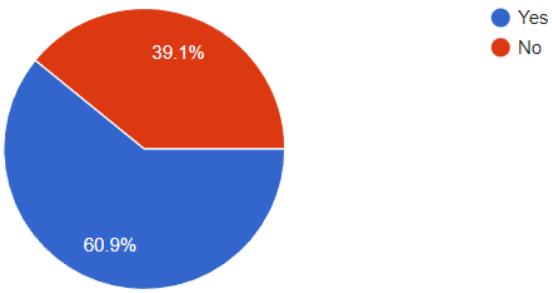
Have you seen these or similar labels while using social media?

23 responses



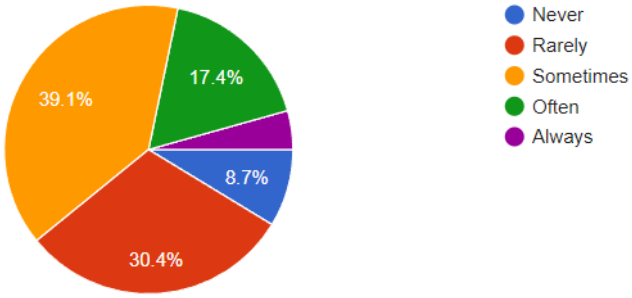
Do you pay for any digital subscriptions of news sources?

23 responses



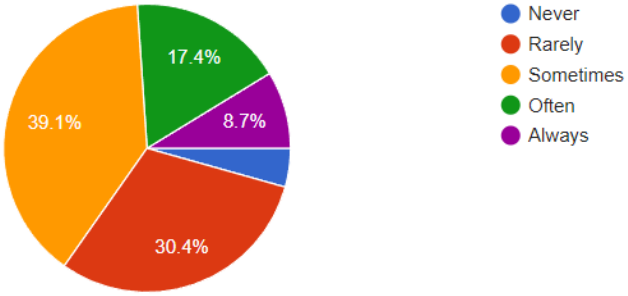
How frequently do you come across fake news?

23 responses



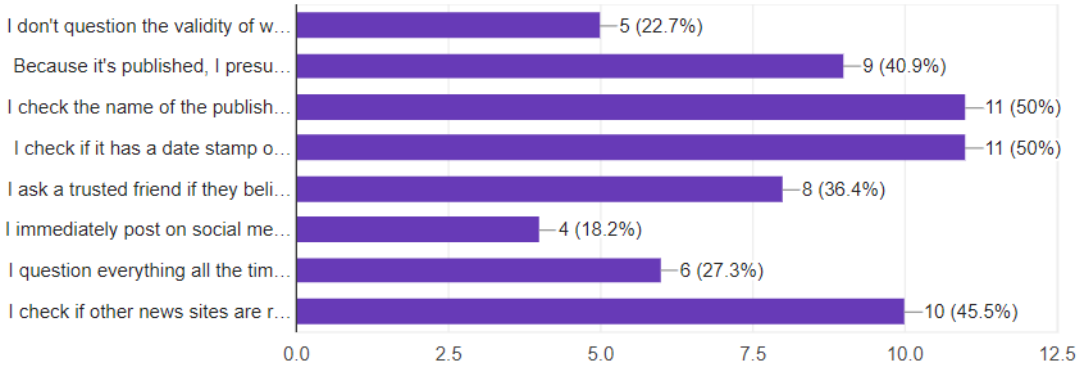
How frequently do you try verify the news you get from sources other than official channels?

23 responses



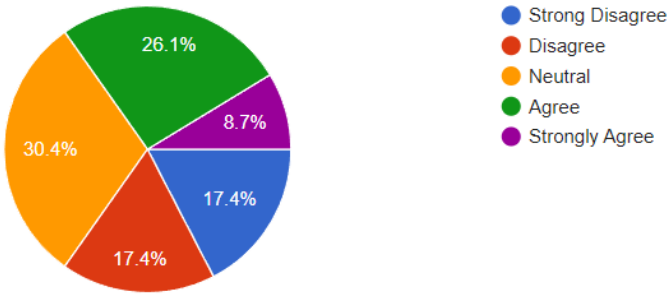
If you do tend to verify news/information, how/when do you do so?

22 responses



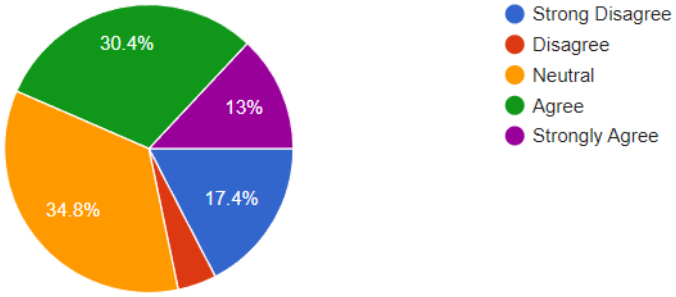
Technology and social media have made me a smarter and more informed person.

23 responses



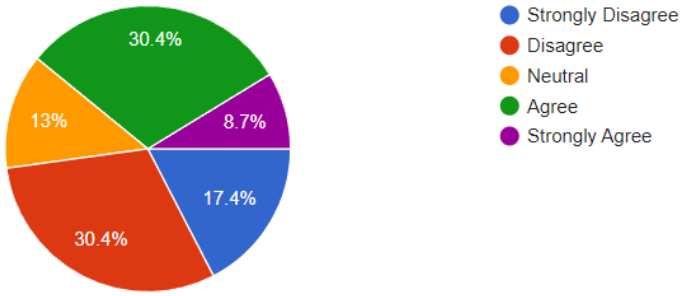
The lack of control and fact-checking on social media makes it suitable for the propagation of unconfirmed and/or incorrect information.

23 responses



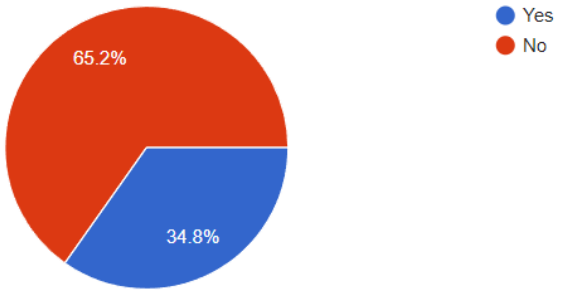
Traditional news outlets don't report fake news.

23 responses



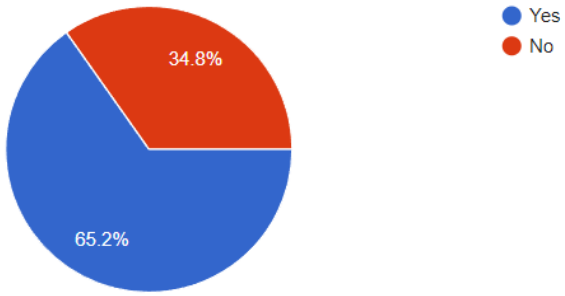
Do you use fact-checking websites like Snopes.com, FactCheck.org, PolitiFact.com and others?

23 responses



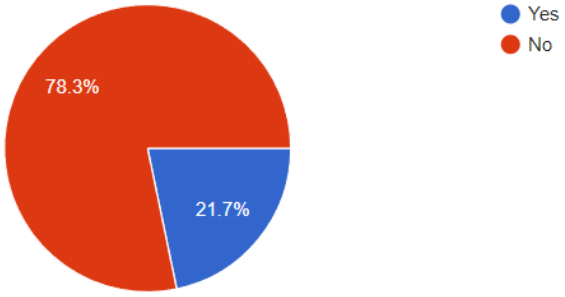
Did your colleagues ever inform you about an article being fake news?

23 responses



Have you ever been affected or seen someone affected due to circulation of fake news in any way?

23 responses



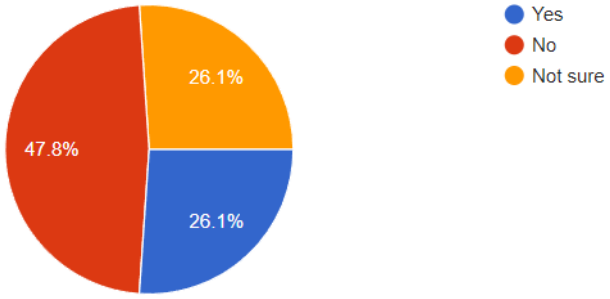
If your answer to the previous question is yes, then how?

1 response

Massive mob lynching in some districts

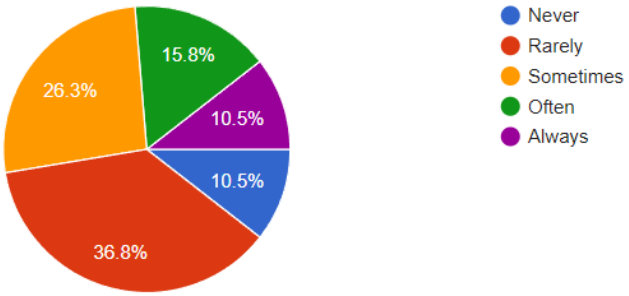
Has any of the news you have shared turned out to be fake ?

23 responses



If your answer is yes for the above question, do you share the corrected version afterwards?

19 responses



Do you have any suggestions regarding confirming authenticity of news?

1 response

Improve reading habits to increase awareness