People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research
Kasdi Merbah Ouargla University
Faculty of New Information and Communication Technologies
Communication
Department of Computing and Information Technology

ACADEMIC MASTER Thesis
Domain: Mathematics and Computer science
Specialty: Fundamental Computer science

Presented by:
Yahia Asma

Theme :

Deep learning-based model for Covid-19 fake news detection

Publicly discussed:

20/06/2022

Before the jury:

Professor Ahmed Korichi	Professor	President	UKM Ouargla
Ms. Toumi Chahrazad	MAA	Supervisor	UKM Ouargla
Mr. Mahjoub Mohammed Bachir	MAA	Examiner	UKMOuargla

Academic year: 2021-2022

Acknowledgement

Above all, I thank Allah who gave me the strength and the will to reach this level and carry out this work.

I thank my supervisor TOUMI Chahrazad for her valuable advice and help throughout this work.

I also would like to thank all the teachers who gave me the benefit of their knowledge and their skills throughout my education.

Finally, I address my gratitude to all those who have contributed, from near or far, to realizing this work by their presence, encouragement and advice.

Dedication

I dedicate this work to:

My dear parents to whom I dedicate all my successes. They have always supported me unconditionally and taught me to work hard for the things that I aspire to achieve.

My brother and my dear sisters for their continuous help and encouragement, without forgetting my little nieces.

The ones who encouraged me to pursue my dreams and supported me till the end.

Asma

Abstract

In the past, we used to get our news from trusted sources, that adhere to specific regulation and editorial standers. With the recent rapid technological advancements of the internet and social media, sharing news is no longer limited to journalists and reporters nor to trusted sources only. Social media platforms have made a revolution in the dissemination of news. They have enabled everyone to consume and deliver news to millions of people around the world in a few seconds of time and almost at no cost. Unfortunately, this rapid distribution of news led to a huge decrease in the quality of the news and it has directly contributed to the wide spread of false information and fake news. Fake news has become prevalent in the age of social media and it has known a huge increase since the beginning of the COVID-19 pandemic. After the COVID-19 outbreak, people started to share millions of posts on social media without considering their reliability and truthfulness. During this pandemic, the spread of fake news and misinformation regarding this virus may cause massive damage and can even place people's lives in risk. Therefore, fighting this infodemic has become a significant challenge for researchers. In this work, we propose a model for detecting fake news shared on social media pertaining to the COVID-19 pandemic. Our model is based on an ensemble of three deep learning models (CNN, LSTM, and GRU). This model was trained and evaluated using the ConstraintAI 2021 COVID19 Fake News dataset which contains news collected from various social media sources such as Instagram, Facebook, Twitter, etc [88]. Our model achieved an accuracy of 93.97% in testing which makes it efficient and capable to evaluate news and make a decision whether they are fake or real.

Key words: : fake news, fake news detection, COVID-19, infodemic, Coronavirus, deep Learning, ensemble learning, Convolutional Neural Networks, Long Short Term Memory, Gated Recurrent Unit.

Résumé

Dans le passé, nous obtenions nos informations de sources fiables, qui adhèrent à une réglementation et à des normes éditoriales spécifiques. Avec les récents progrès technologiques rapides d'Internet et des médias sociaux, le partage d'informations n'est plus limité aux journalistes et aux reporters ni aux seules sources fiables. Les plateformes de médias sociaux ont révolutionné la diffusion de l'information. Ils ont permis à chacun de consommer et de diffuser des informations à des millions de personnes dans le monde en quelques secondes et presque sans frais. Malheureusement, cette diffusion rapide des nouvelles a entraîné une énorme diminution de la qualité des nouvelles et a directement contribué à la large diffusion de fausses informations et de fausses nouvelles. Les fausses nouvelles sont devenues courantes à l'ère des médias sociaux et elles ont connu une énorme augmentation depuis le début de la pandémie de COVID-19. Après l'épidémie de COVID-19, les gens ont commencé à partager des millions de messages sur les réseaux sociaux sans tenir compte de leur fiabilité et de leur véracité. Au cours de cette pandémie, la propagation de fausses nouvelles et de désinformation concernant ce virus peut causer des dégâts massifs et peut même mettre la vie des gens en danger. Par conséquent, lutter contre cette infodémie est devenu un défi important pour les chercheurs. Dans ce travail, nous proposons un modèle de détection des fake news partagées sur les réseaux sociaux concernant la pandémie de COVID-19. Notre modèle est basé sur un ensemble de trois modèles d'apprentissage en profondeur (CNN, LSTM et GRU). Ce modèle a été formé et évalué à l'aide de l'ensemble de données ConstraintAI 2021 COVID19 Fake News qui contient des informations collectées à partir de diverses sources de médias sociaux telles qu'Instagram, Facebook, Twitter, etc. Notre modèle a atteint une précision de 93,55 \ % lors des tests, ce qui le rend efficace et capable d'évaluer les informations et de décider si elles sont fausses ou réelles.

Mots clés: fausses nouvelles, détection de fausses nouvelles, COVID-19, infodémie, Coronavirus, apprentissage profond, apprentissage d'ensemble, réseaux de neurones convolutifs, mémoire à long court terme, unité récurrente fermée.

ملخص

في الماضي ، كنا نتحصل على أخبارنا من مصادر موثوقة تلتزم بقواعد تنظيمية وتحريرية محددة. مع التطورات الإلكترونية السريعة للإنترنت ووسائل التواصل الاجتماعي ، لم تعد مشاركة الأخبار مقتصرة على الصحفيين والمراسلين ولا على المصادر الموثوقة فقط. لقد أحدثت منصات التواصل الاجتماعي ثورة كبيرة في مجال نشر الأخبار. لقد مكنت الجميع من استهلاك ونقل الأخبار إلى ملايين الأشخاص حول العالم في بضع ثوانٍ من الوقت وتقريبا بدون أي تكلفة. لسوء الحظ ، أدى هذا الانتشار السريع للأخبار إلى انخفاض كبير في جودة الأخبار وساهم بشكل مباشر في انتشار المعلومات الخاطئة والأخبار الكاذبة على نطاق واسع. أصبحت الأخبار الكاذبة منتشرة في عصر وسائل التواصل الاجتماعي وشهدت زيادة هائلة منذ بداية جائحة كوفيد-19. بعد تفشي فيروس كورونا كوفيد-19 ، بدأ الناس في مشاركة ملايين المنشورات على وسائل التواصل الاجتماعي دون أي مراعاة لصحتها ومصداقيتها. أثناء هذا الوباء ، قد يتسبب انتشار الأخبار الكاذبة والمعلومات الخاطئة بشأن هذا الفيروس في أضرار جسيمة كما يمكن حتى أن يعرض حياة الناس للخطر. لذلك ، أصبحت مكافحة هذا الوباء تحديًا كبيرًا للباحثين. من خلال هذا العمل ، نقترح نموذجًا للكشف عن الأخبار المزيفة التي يتم مشاركتها على وسائل التواصل الاجتماعي والمتعلقة بوباء كوفيد-19. يعتمد نموذجنا على مجموعة من ثلاثة نماذج للتعلم العميق وهي : CNN، LSTM، و GRU. تم تدريب هذا النموذج وتقييمه باستخدام مجموعة بيانات ConstraintAI 2021 COVID19 Fake News التي تحتوي على أخبار تم جمعها من مصادر وسائل التواصل الاجتماعي المختلفة مثل Instagram و Facebook و Twitter وما إلى ذلك .حقق نموذجنا دقة اختبار 93.55 ٪ مما يجعله فعالاً وقادراً على تقييم الأخبار واتخاذ القرار فيما إذا كانت مزيفة أم حقيقية.

الكلمات المفتاحية: الأخبار الكاذبة ، الكشف عن الأخبار المزيفة ، كوفيد-19 ، الوباء المعلوماتي ، فيروس كورونا ، التعلم العميق ، التعلم الجماعي ، الشبكات العصبية التلافيفية ، الذاكرة طويلة المدى ، الوحدة المتكررة ذات البوابات.

Contents

List of Figures

List of Tables

General Introduction

Indeed, before the internet was invented, people used trusted news sources such as printed media, magazines, radio, and television news to stay informed and keep up with what's going on in their country and in the world. To conduct research, people had to go to the public library and find books, encyclopedias, card catalogs, and other resources to find what they are looking for. Also sharing news to public was restricted to journalists and reporters who adhere to specific editorial standards. Their stories were extensively researched and fact-checked to ensure accuracy and accountability [54][106].

Then, with the advent of the internet and social media, there has been a revolution in the dissemination of news and the concept of news has radically changed. Online news media appeared and enabled everyone to consume and deliver news to millions of people around the world in a few seconds of time and almost at no cost.

Sharing news is no longer limited to journalists and reporters nor to trusted sources only. Online news media, such as blogs and social networks, have made it easy for everyone to create and share news and reports in real time. Since then, there is an increasing pressure to distribute news stories as quickly as possible to avoid being pre-empted by the social media and blogs.

Unfortunately, this rapid distribution of news led to a huge decrease in the quality of the news. It has also resulted in a huge increase in the spread of false information and fake news. It allowed people to have easy and rapid access to information, but it is often difficult to tell whether it is credible or not. The extensive spread of false information can pose major problems and affect society in different ways.

Fake news has become prevalent in the age of social media and it has known a huge increase since the beginning of the COVID-19 pandemic. Commonly known, COVID-19, emerged in Wuhan, China, in December 2019. In the wake of this pandemic, we have witnessed a massive infodemic. An infodemic is Excessive amount of information, including false or misleading information in digital and physical environments during a pandemic. It causes confusion and risk-taking behaviors that can harm health [56]. Social media was being used on a massive scale to keep people safe, informed and aware about the disease and how to protect themselves, yet this technology we rely on to keep connected and informed enabled an infodemic that continues to undermine all efforts and measures to control the pandemic. This infodemic now poses a serious problem for public health [92][58]. Thus, it became vital to create tools that allow us to identify and prevent the spread of fake news about COVID-19.

Quick and early detection of fake news can reduce the spread of panic and confusion among the public. The aim of this thesis is to propose a model for detecting fake news shared on social media pertaining to the COVID-19 pandemic. We use the Contraint@AAAI 2021 Covid-19 Fake news dataset, that consists of news collected from various social media sources such as Instagram, Facebook, Twitter, etc [88]. Our approach is based on assembling three deep learning models: Convolutional Neural Network(CNN), Long Short-Term Memory(LSTM), Gated Recurrent Unit(GRU). This work is divided into three chapters.

The first chapter is divided into two sections. In the first section, we will define false information, its different types, actors and motives behind spreading it. In the second section, we will define in detail fake news, its types, methods of detection and its impacts in the health stage.

In the second chapter, we will define the concepts artificial intelligence, machine learning, deep learning, natural language processing, and ensemble learning. We will also introduce some commonly used deep learning algorithms.

In the third chapter, we will first define the objective and motivations behind this work. Second, we will cite some related works and then we will explain the conception of our proposed model. Next, we will describe in detail the implementation of our model and the dataset used. Finally, we will discuss the results obtained and compare them with other works that used the same dataset.

In the end, we will summarize in a general conclusion the important points discussed in this work, the model proposed for fake news detection, the results obtained, and the future perspectives.

Chapter 1

False information and fake news

1.1 Introduction

In the past, we got our news from trusted sources, that are required to follow strict codes of practice. With the recent rapid technological advancements, social media has enabled the whole world to publish, share and consume information and news easier and with very little regulation and editorial standards. This has contributed to a huge increase in the spread of false information and fake news. It allowed people to have easy and rapid access to information, but it is often difficult to tell whether stories are credible or not. This extensive spread of false information posed major problems and affected society in different ways. Fake news is one of the most common types of false information. In the first section, 1.2, we will talk about false information. We will first explain what is false information, present its different types, actors behind it and their motives. In the second section, 1.3, we will define in detail fake news, its types, methods of detection and its impacts in the health stage.

1.2 False Information

1.2.1 What is False Information?

False information can be defined as incorrect or misleading content, disseminated under the guise of informative content, that contradict or distort conventional understandings of actual facts [89]. Based on the intent behind spreading it, false information is categorized into disinformation and misinformation. Both have negative influences, but the latter can cause more damage as its creator's primary goal is to harm. Disinformation is the creation and dissemination of false information with the intent to mislead, deceive and confuse readers [39]. The intent behind disinformation is to make profit, advertise, harm a person, an organization or a country.

Misinformation is the unintentional dissemination of false information where the person spreading it believes it's true and does not have any malicious intent to deceive [39].

Reasons behind misinformation are multiple we cite:

- The misrepresenting of a true information.
- The misunderstanding of an information by its publisher due to the lack of clarity, knowledge, or attention.
- The cognitive bias, when people tend to believe a specific idea or hypothesis which often gives them an unbalanced and unfair view to the subject.
- Spreading false information for satire.

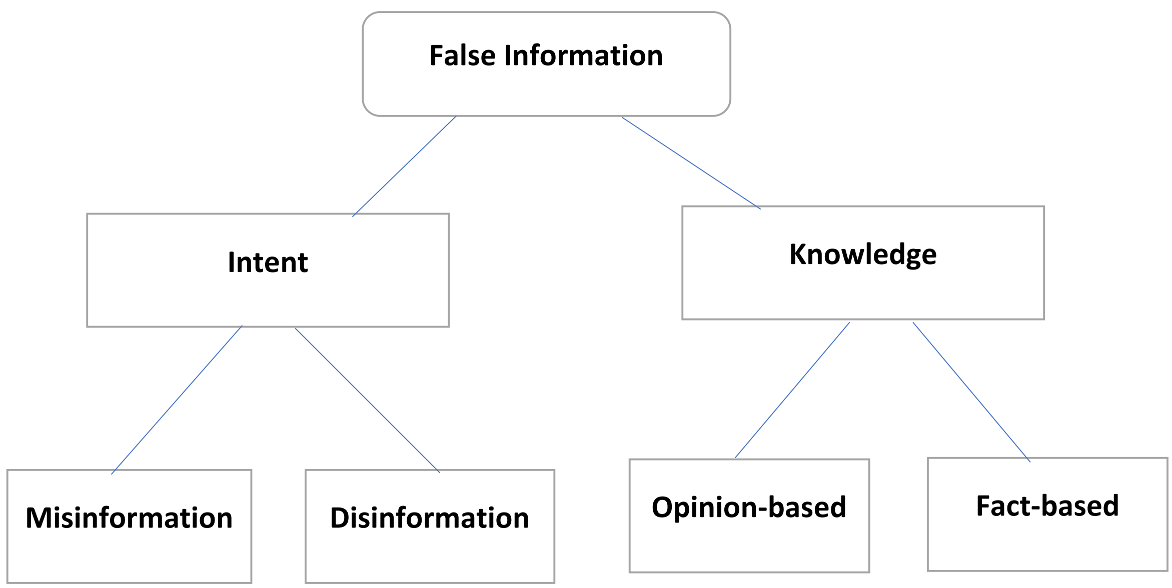


Figure 1.1: Classification of False Information

False information can also be classified based on knowledge into opinion-based and fact-based [67]. Fact-based is relative to spreading an information that contradicts with facts while opinion-based are relative to user personal opinion and point of view. Fact-based information are verifiable facts while in the case of opinion-based, there exist no absolute truth. An opinion can be given dishonestly to effect on others opinions, common example is the applications running in "Play Store" of Android, where developers of applications diffuse some false opinions to deceive people to download.

The intent-based and knowledge-based classification of false information is illustrated in Figure 1.1

1.2.2 Types of False Information

False information exists on the web in different forms. We describe the typology of False Information as: (1) Fake News; (2) Hoaxes; (3) Fake Reviews; (4) Biased or one-sided; (5) Rumors; (6) Clickbaits. The typology of FI is illustrated in Figure 1.2.

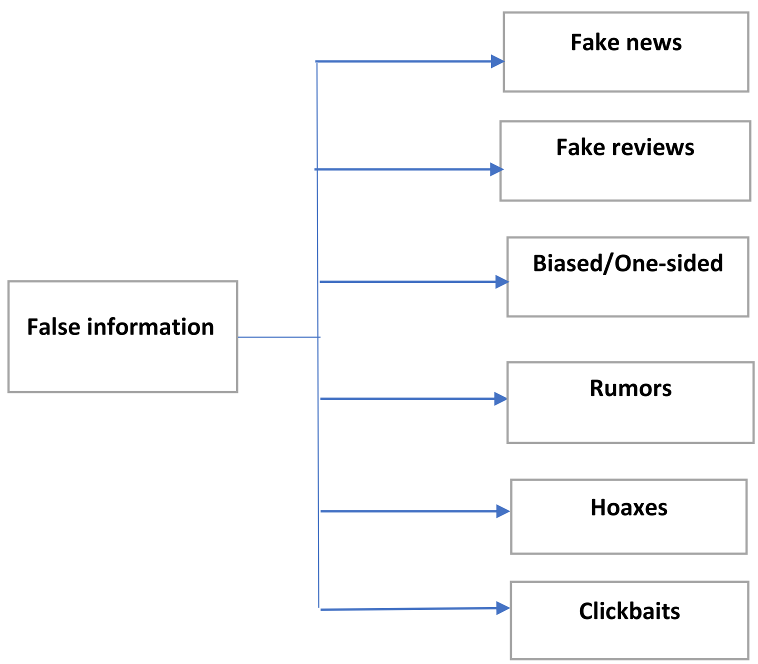


Figure 1.2: Types of False Information

Fake News

Fake News are news that carry wrong and factually incorrect information, they are made up to deceive readers and look like true reports [98]. This phenomenon has existed since the advent of the press, but it became very popular and widespread during the United

States election in 2016 [98]. It is often created for financial or political motives. Users of social media platforms, all over the world read, consume fake news everyday and it is often hard to tell whether what these news are truthful and reliable news or not [40]. We will talk about fake news in detail in Section 1.3.

Fake Reviews

Fake reviews are false, misleading and deceptive online reviews written by individuals based on imagination. Such reviews are inconsistent with real experience and honest evaluations of products or services, they do not reflect the genuinely held opinion of the author [55] [135].
Fake reviews can be positive, negative, or neutral. Different types of people post these reviews online including consumers, online merchants, and review platforms [135]. Fake reviews mislead consumers because they do not reflect the actual product value, they are mainly adopted by online vendors and publishers for boosting a product sales and maximizing the profit or by consumers that are paid and rewarded by the merchant, or sometimes without any financial profit [135].

Biased or one-sided

This kind of news highlight one part of a story to extremely favor a particular person, party or situation. Biased news media and reports have viewers with inaccurate, unfair, unbalanced and emotional view of the world around them which influences their opinions and choices. This type of false information is usually found in political conflicts known as Hyperpartisan news [94] and feeding the public by only one side of the issue. Almost everyone's social network account contains bias stem from the social interactions of their accounts [93].

Rumors

Rumors are circulating information that are uncertain and their truthfulness was not verified at the time of posting. This ambiguous information circulates on social media platforms and eventually it might turn out to be true, false , partially false or never confirmed [141]. This type of false information spread very fast in social networks due to the high flexibility they provide in sharing and exchanging information. This has made these platforms a fertile ground for rumormongers to post and spread rumors; it can cause unpredictable reactions from involved individuals and real-life incidents of damage and chaos during breaking news [5].

Hoaxes

Hoaxes are popular type of false information, can be defined as news stories that contradict the facts. These stories contain false or inaccurate news presented as legitimate facts [137]. This definition applies only to information that has no basis of truth [80].

Popular examples of hoaxes are stories that report the false death of celebrities like the hoax spread in 2012 about the death of Algerian president Bouteflika [119].

Clickbaits

Clickbaits refer to the use of misleading headlines and descriptions of content in order to attract attention and encourage visitors to click a link to a particular web page. It is often created for financial gain and it is one of the less harmful kinds of false information because it can be easily detected after clicking the link and reading the content [137].

1.2.3 False Information Actors

The actors of false information are the creators who are responsible of creating the false information or spreaders who share the false information and help in its propagation. There exist many types of actors creating and spreading false information on the web, they are listed below:

Political Organizations

Many politicians tend to diffuse false information in order to attract the audience to their organizations or discredit other organizations. We find this type of actors very active in the periods of elections. Sometimes even Governments are involved in the spreading of false information to manipulate their public opinion on specific topics or even sow discord in other countries. The US. elections in 2016 are very popular example of fake news shared by political organizations. Facebook was accused for disseminating false information and that affected the results of the elections [137]. In the same context, according to the FireEye [1] researchers, suspected Russian operators used bots to spread anti-Clinton messages in Twitter and Facebook and bought ads pushing divisive issues during and after the American election campaign [108].

Journalists

Journalists are entities responsible for carrying the information to the public. In some cases, they may publish some false information. This can be due to reckless publishing before verifying the information or deliberately publishing it either for profit or for other malicious intents.

Trolls

Trolls are real accounts who deliberate spreading false and controversial information to incite others, create emotional strain among network's users [137], and lure them into fruitless argumentation [26]. Trolls intend to exploit "hot-button issues" to attract users

[1] www.fireeye.fr

to react emotionally [19] and provoke them to engage into arguments and debates in order to create disagreement and conflict [26] or just to amuse the interactions of users [19].

Bots

Online fake accounts that are predominantly controlled by software rather than a human user. Bots are created to mimic real users(humans) on a given platform and spread a story to the maximum sum of people. They are generally controlled by a single underlying entity [67] for malicious acts like spam, propagating fake news and rumors, and making noise more frequently than legitimate usage.

Sockpuppets

The term ”sockpuppet” refers to the manipulation of a simple hand puppet made from a sock [113]. On virtual spaces, some individuals use false identity to communicate with others. They forge other users and create content in which they express themselves pretending to be another person [18]. Sockpuppets aim usually to create fake support for a product or one's work. They are used for business promotion, product review, and political influence. Clemson University social media researcher identified in October of 2020 ”more than two dozen of Twitter accounts claiming to be black Trump supporters who gained hundreds of thousands of likes and retweets in a span of just a few days, sparking major doubts about their identities” [113].

Hidden Paid Posters

The term hidden paid posters is referred to as the “internet water army” in China because of the big number of people organized to “flood” the internet with purposeful comments and articles [22]. Paid posters represent a new type of online job opportunity on the internet. For many online users, particularly college students and the unemployed people, this job is a valuable opportunity. These users get paid for posting comments and new articles or content on many online communities and websites for some hidden goals like influencing the view of other people towards certain entities, events, services, or products [22]. For example a company that hire some users to spread wrong and negative information about its competitors is using hidden paid posters. This kind of actors is similar to bots but they are harder to distinguish because they have characteristics of regular users.

Individuals that benefit from false information

Individuals spreading some false content for their personal gain and motives like promoting their services and attracting costumers.

“True Believers” and ”useful idiots”

People sharing false information while believing that they are sharing the truth and informing people of unknown facts. This kind of spreaders are often passionate about an idea or an entity, which affects their judgement to the relative subjects. They, for example, might take a side of an ideological debate without being aware of the malicious agenda driving the ideology [13] .

Criminal/Terrorist Organizations

Criminal organizations tend to use social media to misleading public opinion and deceive people to unconsciously cooperate in achieving their criminal goals. According to [11], Several fake accounts managed in Algeria were accused to be involved in the case of the young man murdered Djamal Ben-Smail, many efforts and operations to spread false information on Facebook and Twitter were organized ahead of the young man’s death.

1.2.4 Motives behind false information propagation

Motives behind spreading false information are various, some actors do it for just fun, while others have malicious intent and desire to harm a person or an organization by hurting their image and reputation. Meanwhile, others can spread false information to influence and manipulate people’s decisions and public opinion in specific subjects. The aim behind influencing people can be collecting followers, changing some norms in the community, imposing a political opinion, etc. Another motive for propagating false content is to sow discord and confusion among people. This can appear in the political stage, like the Russian efforts to sow discord in America [12]. Ranking Member Adam Schiff stated, on November 2017, about the Russian social media campaign: “social media campaign was designed to further a broader Kremlin objective: sowing discord in the U.S. by inflaming passions on a range of divisive issues. The Russians did so by weaving together fake accounts, pages, and communities to push politicized content and videos, and to mobilize real Americans to sign online petitions and join rallies and protests” [12].
From another side, some individuals may have some passion towards an idea or an entity that they unintentionally contribute in spreading false information about it. This kind of people are often blinded by their ideology. And finally, the most common motive behind spreading false information is profit. A lot of individuals and organizations tend to increase their profit by disseminating false information. They seek for views and comments by manipulating users and deceiving them. A common example used in Facebook, where pages diffuse some good products for cheap prices, users get attracted and interact with the post by commenting and sharing it, the page gains popularity due to the views and comments and increases its profits. Another way for raising the number of views is using clickbait. Manipulators want to gain money by diffusing interesting headlines to attract people to click and open their link [137].

1.2.5 Sources of False Information

False information is widespread on the web, from its multiple sources we cite:

Social Networks

The biggest source of false information, they are sites that connect people with each other and allow them to share stories publicly or privately. Some social networks have social purpose like Messenger, some have a business purpose like LinkedIn and other have both like Facebook. [63].

Q&A platforms

A Q &A platform in an online software that attempts to answer questions asked by users. Q&A software is a community that enables users in similar fields to discuss questions and provide answers to common and specialist questions and it is frequently integrated by large and specialist corporations [97]. One of the most famous Q&A platforms is: Stack Overflow [2].

Review platforms

Review platforms are websites on which people can leave reviews about businesses, products, entities, or services. These sites may engage professional writers and authors to produce reviews on the site's area of focus, or they may use specific strategies to collect reviews from site visitors [101]. Some review platforms are: Yelp[3], and Trip Advisor[4].

Video sharing platforms

Video sharing platforms are platforms that enable people to upload and share their video clips with the public at large or to specific guests. There exist many video sharing platforms like: Youtube [5], Vimeo [6], DailyMotion [7], and TikTok [8].

Blogs

Blogs are online personal-journals where the owner(s) presents information or share his views on a specific domain. The blogs are frequently updated and consist of dated entries called posts. Posts are displayed in reverse chronological order, with the latest post appearing first, at the top [74]. Each post can have three basic attributes: title, link, and

[2] stackoverflow.com
[3] yelp.com
[4] tripadvisor.com
[5] www.youtube.com
[6] vimeo.com/fr
[7] www.dailymotion.com
[8] www.tiktok.com/fr

description(text content). This type of websites is being used since the development of the World Wide Web (WWW) [17]. Examples of famous blogs for news [78]: BuzzFeed [9], ThinkProgress [10], Mashable [11] and The daily Beast [12].

Micro-Blogs

Micro-blogging is the practice of sharing short messages or posts with an online audience using microblogging services. It is a type of blogging that allows people to broadcast their thoughts online. Micro-blogs differ from regular blogs in that the information is often brief and lower in size [75][65]. Microblogs enable users to exchange small content such as short phrases, isolated photos, or video [75]. Many micro-blogging platforms exist like: Twitter[13], Instagram[14], and Tumblr[15].

Forum Sites

Websites that allows its users to communicate with each other and exchange information and ideas on topics of interest by posting and adding comments. Most forums often allow anonymous visitors to view forum posts, but require registration for posting or commenting in the forum [130]. Here are two examples of famous forums: Reddit [16], Quora [17].

News websites

Online Newspapers devoted to delivering news in a digital form, they can be versions of printed papers. Unlike printed newspapers that require waiting for the next day to publish their news, these sites provide immediate broadcast of the latest and breaking news. These sites may include all types of news like political, judicial, entertainment, sports or more. Examples of this type of sites: The Guardian [18], NBC News [19] and CNN news [20].

Social news websites

Websites characterized by user-posted stories, these communities encourage users to publish and share news stories, articles and media (images/videos). They depend on the "wisdom of the crowds" principle where content disseminated through the community will

[9] www.buzzfeed.com
[10] thinkprogress.org
[11] mashable.com
[12] www.thedailybeast.com
[13] twitter.com
[14] instagram.com
[15] www.tumblr.com
[16] www.reddit.com
[17] fr.quora.com
[18] www.theguardian.com
[19] www.nbcnews.com
[20] edition.cnn.com

be ranked based on popularity and the value or importance will be determined based on the interactions of users [132].Digg [21] is a very popular example of a social news website.

Fake News websites

Websites that publish wrong and misleading news, purporting to be real news, intentionally to defraud readers and drive web traffic. Unlike news satire, fake news websites deliberately seek to be perceived as legitimate and taken at face value, often for financial or political gain. Some of fake news websites use website spoofing, they impersonate other trusted websites adopting their design and taking similar URLs. These spoofing websites aim to make visitors believe that the website is created by another person or organization and they are visiting trusted sources [38]. Breaking-CNN [22] is an example of fake news websites, it is designed to emulate CNN [23] and it is responsible for publishing numerous death hoaxes, including one for former First Lady Barbara Bush in 2018 [71].

Satirical websites

News satire are popular on the web, they are published in satirical websites like TheOnion [24]. These websites present news using irony and humor. Differing from fake news websites, the purpose of satire news websites is not to deceive but to deliver joy and cheer the audience and audience are aware that news presented are not always correct [116].

Collaborative Projects

Collaborative projects allow teams of two or more participants with differing inputs to work together and create online content. They equip us to manage all our projects, teams, clients, documents, and stakeholders using a single platform and they help us to achieve high-efficiency levels by making project management seamless. Some popular collaborative projects websites are: GitHub [25] and Wikipedia [26].

1.3 Fake news

1.3.1 What is Fake news?

The term ***News*** appeared very long time ago. In the past news were created and distributed in conventional news media like news papers, magazines, radio, and television as a source of information. Sharing news was not allowed to everyone and they were often

[21] digg.com
[22] Breaking-CNN.com
[23] edition.cnn.com
[24] www.theonion.com
[25] github.com
[26] wikipedia.org

generated by journalists. With the advent of internet, the concept of news has radically changed. Online news media appeared such as: news websites, blogs, and social media. However, creating and sharing news and reports became easy and accessible to everyone. This rapid distribution of news led to a wide spread of false information; People started spreading fake news widely for different motives and purposes.

According to [139], "The definition of news broadly includes articles, claims, statements, speeches, and posts, among other types of information, related to public figures and organizations. It can be created by journalists and non-journalists."

Fake news have become a major phenomenon in the context of social media. It represents false and incorrect information presented as news article or message, propagated through online mainstream media and designed to mislead readers and deceive them into believing they are consuming real news [6] [139] [44].
Fake news is a very old concept but it was widely spread after the US. elections in 2016. During the critical months of election, Twitter knew an expansion in fake news sharing and this term gained a lot of attention since [6].

Fake news is characterized by imitating news media and pretending to be real. They attempt to appear like real news by taking a form of credibility and legitimacy. In addition to the news' form, fake news also use fake accounts and bots to propagate in the network. The motives behind the creation and dissemination of fake news are various (often financial or political).

1.3.2 Types of fake news

Fake news can appear in many forms and since fake news is a type of false information, these forms can be considered as false information as well. We list the different types of Fake news below:

Satire News

Refers to TV shows or news articles that present new updates and stories using irony and humor [116]. Differing from fake news, the purpose of satire news is not to deceive but to make fun, ridicule and cheer the audience, the main goal is to deliver joy and entertainment rather than information. It is mostly created by entertainers not journalists and audience are aware that news presented are not always correct [61] [127]. Satire news uses the style of a television news broadcast and often focuses on current affairs and uses humor to provide critique of political, economic or social affairs. [116]. Sometimes, audience with no knowledge might mistakenly consider news satire as actual news [127], therefore satirical websites often disclose their satirical nature in the description provided by their website [137].

Parody News

Parody, like satire, uses humor to draw its audience who is aware that the information presented are fictitious, but the main difference is that the news stories used to inject humor are entirely fictional [98] [116].

Advertorial News

Advertorial, a compound term combining "advertising" and "editorial". Advertorial news refers to commercial messages and advertising materials contained in news. Mainly these advertisements are paid by a sponsor, they can mislead audience by inserting marketing or other sponsor messages into news media that adopt the legitimacy of the news format, by focusing only on the positive aspects of the advertised product or company and by pretending that the news produced is free of bias. [61] [116].

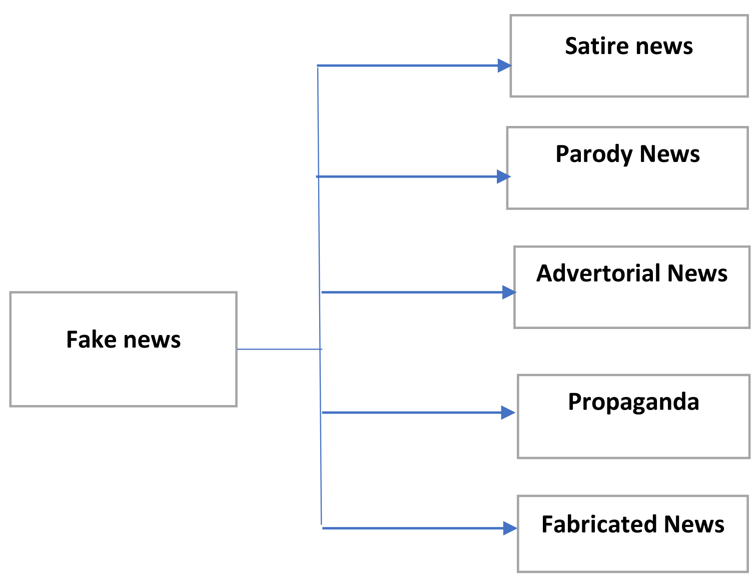


Figure 1.3: Types of Fake News

Propaganda

Propaganda is the deliberate attempt of a political entity to influence perceptions and behaviors, the propagandist promote particular thoughts and ideas to further their goals [28]. Propaganda was widely used during WorldWar II and the ColdWar [137]. It is an instance of fake news, very used in political contexts and aims to present a story for a favorable side that could harm the interests of other parties and insert bias. Very popular terms are related to propaganda like news management and spin. Spin is often used with reference to the manipulation of political information . An example of propaganda in 2002, when the United States launched a propaganda assault on Saddam Hussein's subjects and soldiers, with a radio station and a massive drop of leaflets this week. Several attempts were made trying to dissuade the Iraqi military from supporting Saddam [36].

Fabricated News

refers to news that have no factual basis but are presented in the style of news articles to give them credibility and legitimacy. Unlike parody, the intention of authors is misinforming. Readers are not aware that the news presented are false, they believe they are consuming real news.

The typology of FN is illustrated in Figure 1.3.

1.3.3 Fake news detection

Social media represents the main source of fake news [68], users are daily consuming news on these platforms due to their low cost and the easy and rapid access to the information [110]. The wide spread of fake news in social media and its negative impacts on society led to tremendous interest in this field, fake news detection became a challenging problem and attracted the attention of researchers.
Fake news detection is the process of making prediction whether a news article is real or fake. There exist different approaches in literature to detect fake news in online platforms. We categorize existing methods of fake news detection based on their main input sources as: News Content methods and Social Context methods. as follow:

Content based methods:

Aim to detect fake news by analysing the information and features provided in the news content such as: news text, headline, image, video, and source. To detect fake news based on content the following approaches can be used:

- **Knowledge-based methods** Knowledge-based approaches are used to check the truthfulness of major claims in a news article to infer its veracity [110] [72] [139]. These approaches use external sources to verify the authenticity of a given claim, this process is called "fact-checking".
  Fact-checking aims to identify whether a claim in a specific context is true and factual

by comparing it with known facts from external sources [126]. This process can be done manually or automatically.

- **Manual Fact-Checking** relies on human judgement to verify the truthfulness of a claims. It is often conducted by expert fact-checkers or news organisations and journalists [79]. Well known fact checking websites are: Reporterslab [27], Polifact[28], The Washington Post Fact Checker [29], FactCheck [30], Snopes [31], TruthOrFiction [32], FullFact [33].
  Reaserchers rely on these websites in creating and labeling their datasets [50]. The manual fact checking can be realised by experts or crowd-sourcing [50].
  - **Expert based** Relies on human experts in specific domains to judge the veracity of relevant claims and data. Websites like Polifact[34] and Snopes [35] use this approach.This process is effective and gives highly accurate results but it has many limitation: it is costly, time consuming and hard to scale well with the big volume of news being created every day therefor, automatic fact-checking has been developed as an alternative to detect the veracity of the news [139] [100].
  - **Crowd-sourcing sourced** This approach exploit the "wisdom of crowd" [50], it enables a large population of individuals to evaluate the veracity of claims and act as fact-checkers [139]. Compared to expert based , crowd sourced fact checking have better scalability but it is less credible and accurate and may contain conflicting annotations because fact-checkers have political and ideological bias, therefore the fact-checkers need to be tested and filtered beforehand and the conflicts need to be resolved later [139]. For example: the crowd sourcing marketplace Amazon Mechanical Turk (AMT) was utilized in the annotation of fake news dataset CREDBANK [77]. Fiskkit [36], as well, allows users to discuss and find what information is accurate in a news article.
- **Automatic Fact-Checking** As the manual fact checking is not scalable with the big volume of created data, automatic techniques have been developed to perform this task in a reasonable time, these techniques rely mostly on artificial intelligence. This process can be divided into two sequential phases:

[27] reporterslab.org
[28] http://www.politifact.com
[29] www.washingtonpost.com/news/fact-checker
[30] www.factcheck.org
[31] www.snopes.com
[32] www.truthorfiction.com
[33] fullfact.org
[34] http://www.politifact.com
[35] www.snopes.com
[36] fiskkit.com

fact extraction then fact checking. In fact extraction, knowledge is extracted from reliable sources(one source or many sources), then in the fact checking, the data to be checked is compared with the knowledge extracted and then classified as factual or non-factual [139].

- **Style-based methods** While claims are generally verified by checking the authenticity of the information, there are other characteristics that can be extracted from the writing style of the news that help to identify the deception in the text. Instead of the authenticity, style based techniques predict the intention of the writer to mislead by analysing the style of the written news [139]. The goal of style-based approaches is to capture the difference in the characteristics of the writing style between the legitimate and the deceptive content. A study made by [53] proved that fake and real news articles differ in the writing style where fake news titles are longer, use few stop words, fewer common nouns and more proper nouns, more all-capitalized words, more verb phrases, more past tense words and put the main claim in the title which is often about a specific person or entity, whereas the body of the article tend to be short, repetitive and less informative.

- **Visual-based methods** Visual-based methods use visual features that are represented in images and videos attached with the news article, these features are often are used as cues to frame the story, evidence to prove it or to attract attention. Misappropriation and misrepresentation of visual content like images help the growth of fake news because people tend to see images as truthful evidence. In a study made by [59], the accuracy of fake news detection, taking image content in consideration, increased by more than 7% compared to the approaches using only textual features. Image features were taken from two aspects: visual content and statistics. Whereas in [49], a classification algorithm was developed to distinguish between tweets containing real and tweets containing fake images. Two kinds of features were used: user level and tweet level features. However, most existing approaches of fake news detection in literature rely on textual features extracted from the news text and ignore the visual features.

Social context-based

Compared to content-based, there exist only few works on social context-based detection of fake news in literature [110]. However, content-based techniques are not sufficient because fake news are written intentionally to look like true; therefore, context based-features can be used to add auxiliary information to infer the deception. These techniques focus on analysing the users' social engagements features [72]. Users usually tend to build connections based on their common interests, these connections act like channels that disseminate information on social media.

Social context models basically include stance-based and propagation-based. Stance-based approaches highlight users' opinions and viewpoints from relevant post contents to detect the veracity of original news. The stance of users' posts can be represented either explicitly or implicitly [110]. Propagation-based models focus on the interrelations of relevant social media posts to predict news veracity. The credibility of a news event is highly related to the credibility of relevant social media posts [110].

Hybrid methods

Content-based methods are effective and widely utilized in fake news detection, hence users' social engagements and reactions with news on social media provide abundant auxiliary information for better detection [72]. This hybrid approach emerged as an amelioration and solution to the restrictions present when using only one of the methods. [24] shows in their work that combining content-based and context-based methods may lead to better and more accurate results. Further, [109] propose an hybrid approach and prove its effectiveness by experiments.

1.3.4 Impacts of fake news in health

With the increasing amount of false information available online, people might quickly be deceived. This issue is particularly relevant when misinformation is related to health. People can have an active role in consuming and evaluating health information on the web. However, laypeople are not health experts and it will be difficult to predict the reliability of information they read; this may cause negative impacts to the ones who are exposed to misinformation spread especially during pandemics or those researching their own ailments online.

Misinformation about the pandemics presents a serious threat to public health and public action. The impact of fake news in social media is a serious public health issue, since it has the potential to increase or decrease the success of programs, campaigns, and initiatives aiming at improving people' health, awareness, and well-being.

After the outbreak of coronavirus (COVID-19) in 2019, millions of fake posts and articles were disseminated online about this virus causing massive impacts. There is much concern about the dissemination of misleading or inaccurate information during this pandemic. In February 2020, the World Health Organization (WHO) warned that the COVID-19 outbreak had been accompanied by a massive 'infodemic', making it difficult for people to find reliable sources and trustworthy information when they needed it. Social media usage, the lack and the low level of health/eHealth literacy are identified as the major causes of this infodemic. This latter caused a lot of panic and public psychological issues like: uncertainty, fear, anxiety, racism loneliness, and depression on a scale not seen in previous epidemics, such as SARS, MERS and Zika [95][92]. It also caused the inappropriate protective measures applied by people believing instructions found on social media and the loss of trust in health institutions and programs [92]. Among the conspiracies shared by social media users in this period, a popular theory has linked 5G to the spread of COVID-19 leading to misinformation and the burning of 5G towers in the United Kingdom [3]. Another theory suggested that COVID-19 had been manufactured by international authorities and agencies as a bio-weapon [58]. Multiple countries were also accused of manufacturing and spreading the deadly coronavirus in China as part of an economic and psychological campaign against the country [58]. Others, on the other hand, alleged that the virus was created in a laboratory as part of China's bio-warfare program26 and that a Chinese scientist had built it as a weapon[58]. There have also been reports of death related to this infodemic. A man in India committed suicide after mistakenly believing he had coronavirus infection. According to the deceased's family members, the person felt guilty and ashamed for contracting COVID-19, he was afraid that the virus had unknowingly spread to his family members, and he was anxious of how the society would react [58].

1.4 Conclusion

We divided this chapter into two main sections. In the first one, 1.2, we have explained what is false information, and we have presented its different types, actors behind it and their motives.

In the second section, 1.3, we have first defined fake news in detail. Then we have introduced its types and methods of detection. In the end we have presented its different impacts in the health stage.

In the next chapter, we will talk about machine learning and deep learning that are very efficient tools for the automatic detection of fake news.

Chapter 2

Machine Learning and Deep Learning

2.1 Introduction

The widespread dissemination of fake news on social media poses significant negative effects on many levels, and predicting the truthfulness of claims became a challenging task even for humans. To tackle these challenges, computer scientists are making efforts to develop solutions for identifying and mitigating fake news with the help of machine learning. Machine learning is a form of artificial intelligence. It focuses on creating systems that learn and improve their performance based on the data they process without being explicitly programmed by human. Deep learning, derived from machine learning, is a promising alternative to traditional machine learning methods. It shows excellent performance for large datasets in different tasks including natural language processing and text classification. For better results, multiple deep learning models can be combined to perform together using ensemble learning.
In the first section of this chapter, 2.2, we will introduce the concepts of Artificial Intelligence and Natural Language Processing. Then, in Section 2.3, we will talk about machine learning, its different types, algorithms, and applications. Next, in Section 2.4, we will present what is Deep Learning and then define some of its popular algorithms in Section 2.5. Finally, we will define ensemble learning and its types in Section 2.6.

2.2 What is Artificial Intelligence?

Artificial intelligence (AI) is a branch of computer science. It involves developing computer programs to complete tasks which would otherwise require human intelligence [9]. AI enables computers to mimic human behavior and it is widely used within the modern world, such as personal assistants, self driving vehicles or chess playing programs.
AI is evolving very quickly, its main domains of application are: machine learning (ML), natural language processing (NLP), language synthesis, computer vision, robotics, sensor analysis, optimization and simulation, and expert systems.

2.2.1 What is Natural Language Processing?

Natural Language Processing (NLP) is the automatic manipulation of natural language, like speech and text, by software [15]. It is a branch of artificial intelligence that enables computers to analyze, understand, or produce written or spoken human language.
Indeed, a typical computer understands a very precise, structured and unambiguous programming language, but the human language is imprecise and confusing. NLP allows the machine to comprehend human language and identify the important features from it.

2.3 What is Machine Learning?

Machine Learning (ML) is a subset of AI techniques. It enables computer systems to learn from previous experience (i.e. data observations) and improve their behaviour for a given task without being explicitly programmed by a human [85].

Machine Learning systems have many applications like identifying objects in images, transcribing speech into text, match news items and products with users' interests, or select relevant results of search.

2.3.1 Types of Machine Learning

Supervised Learning

A supervised learning algorithm generates a function that maps inputs to desire outputs [138]. This type of machine learning requires a labeled dataset where the algorithm trains and learns some rules from the data provided with the labels and generates a trained model that is able to predict the classes of new instances. This generally requires a training set for training the model and a testing set for evaluating its performance. Supervised Learning can be used in classification problems or regression problems. In classification, the labels(classes) to predict are discrete values, the algorithm aims to predict to which category or class the instance belongs. While in regression, the outputs generated from learning are continuous values. A very common example in supervised classification is the digit recognition problem. The supervised classification algorithm uses the dataset of handwritten digit pictures provided with classes(the digit corresponding for each picture) and learns the features and proprieties of each digit in order to generate a classifier that is able to recognize the new handwritten digit data.
Some algorithms used in supervised learning are: K-Nearest Neighbor [90], Naive Bayes [131], Decision Trees [103], Linear Regression [115], Support Vector Machines (SVM)[73], Random Forest [102], Neural Networks [1], etc.

Unsupervised Learning

Unsupervised Learning also involves training of the data but with no labels. The machine tries to cluster the similar types of the data and aims to discover the patterns rather than making predictions. The evaluation of the model in unsupervised learning cannot be done as the data is not labeled [31]. The algorithms involved in unsupervised learning are K-mean clustering [136], Association Rule Mining [66], Topic Modeling [125], Dimensionality Reduction Techniques [114], etc.

Semi-supervised Learning

As supervised learning works on labeled data and unsupervised learning on unlabeled data, then a lot of information is lost from labeled data which can be obtained from unlabeled data. So, in this case semi-supervised learning comes to mind. It is a mixture of supervised and unsupervised learning in which it takes both the unlabeled and labeled data. Labeled data should be of shorter length as compared to unlabeled data. The idea behind semi-supervised learning is that there is a considerable change in performance when both labeled and unlabeled data is used in conjunction. The training set used is of shorter length. It is normally used to detect outliers [31].
Many algorithms can be used in unsupervised learning such as k-means clustering [112] or Association Rules [66].

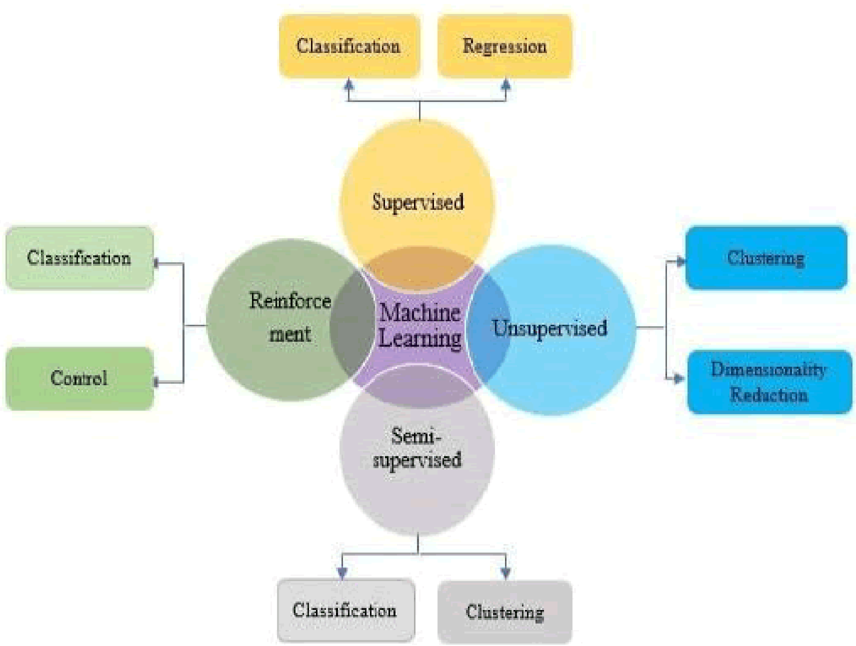


Figure 2.1: Types of Machine Learning [31]

Reinforcement Learning

Reinforcement Learning is a machine learning training method that works by developing a system which improves its performance by taking feedback from the environment [31]. It is an iterative process based on rewarding desired behaviors and punishing undesired ones. In general, a reinforcement learning agent perceives and interprets its environment, takes actions, and learns through trial and error. The common algorithms used in Reinforcement Learning are: Q-Learning [129], Temporal Difference (TD)[118], and Deep Adversarial Networks [45].

The different types of machine learning algorithms and their applications are shown in figure 2.1.

2.3.2 Algorithms of Machine Learning

Below is the list of most commonly used Machine Learning algorithms:

Linear regression

Linear regression is a linear approach that provides a mathematical formula which is able to generate predictions. It models a linear relationship between the input variables (x) and the single output variable (y). It means that y can be calculated from a linear combination of the input variables (x). When there is one single input variable (x), the method is called simple linear regression; for more than one, it is called multiple linear regression. Learning a linear regression model entails estimating the values of the coefficients in the representation using the available data. To prepare or train the linear regression equation from data, different techniques can be used like: Simple Linear Regression, Ordinary Least Squares, Gradient Descent, Regularization. The most common method used is Ordinary Least Squares. It is common to therefore refer to a model prepared this way as Ordinary Least Squares Linear Regression or just Least Squares Regression. Linear regression can be used in various areas like commerce and business to make predictions for better decisions [70][69].

Logistic Regression

Logistic Regression [27] is a powerful supervised ML algorithm for classification problems, it is useful for predicting the category into which the sample fits best. From a set of independent variables, logistic regression is used to estimate discrete values (usually binary values like 0/1). By fitting data to a logit function, it aids in predicting the probability of an event. It's also referred to as logit regression. Logistic regression can achieve good performance in attack detection.

Decision Tree

Decision Tree is one of the most used machine learning algorithms nowadays [120], it uses a tree-like model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility [30]. Decision trees are supervised learning algorithms for classifying problems. They are effective in classifying both categorical and continuous dependent variables [120]. Decision trees are mainly used in operations research, specifically in decision analysis, to help discover the best method for achieving a goal [30].

Support vector machine

Support vector machine (SVM) is a machine learning algorithm that can be used for both classification or regression challenges [99]. It is a method in which raw data is plotted as points in an n-dimensional space (where n is the number of features you have). Each feature's value is subsequently linked to a specific coordinate, making data classification simple. Classifiers are lines that can be used to separate data and plot it on a graph [120].

Naive Bayes Algorithm

Naive Bayes are a family of simple "probabilistic classifiers" based on applying Bayes' theorem with strong (naive) independence assumptions between features [81]. The existence of one feature in a class is assumed to be independent to the presence of any other feature by a Naive Bayes classifier. Even if these characteristics are connected, a Naive Bayes classifier would analyze each of them separately when determining the likelihood of a specific outcome [120]. A Naive Bayesian model is simple to construct and can be used to analyze large datasets. It's easy to use and has been shown to outperform even the most complex categorization systems [120].

K Means

K Means is a clustering problem-solving unsupervised learning algorithm. Data sets are divided into a certain number of clusters (K) in such a way that all data points within each cluster are homogeneous and distinct from data in other clusters. In order to form the clusters, the K-means algorithm selects k centroids, or points, for each cluster. With the closest centroids(K clusters), each data point forms a cluster. It now generates new centroids based on the members of the existing cluster. The closest distance between each data point is determined using these new centroids. This cycle is repeated until the centroids remain stabilized (there is no change in their values) which means that the clustering has been successful, or the specified number of iterations has been completed [123][120].

K Nearest Neighbors

K Nearest Neighbors (KNN) is a non-parametric, supervised learning classifier that makes classifications or predictions about the grouping of a single data point based on proximity [60]. It's a simple algorithm that saves all existing examples and classifies any new ones based on the votes of its k neighbors. The case is then placed to the class that shares the most similarities with it. This measurement is performed via a distance function. This approach can be used to solve both classification and regression problems [120].

Random Forest

Random forest, like its name implies, consists of a large number of individual decision trees that operate as an ensemble. Each tree in the random forest produces a class prediction and votes for it. The class with the highest votes is chosen by the forest as the prediction of the model [120][124]. Ensemble predictions produced by uncorrelated models can be more accurate than individual predictions because the trees protect each other from their individual mistakes. While some trees may be incorrect, many others will be correct, allowing the trees to move in the correct direction as a group [124].

2.3.3 Applications of Machine Learning

Machine Learning powers many aspects in our modern society like:

- **Image recognition** distinguishing a specific object or objects in a picture like identifying faces for security application, detect traffic lights for self driving vehicles, detect elements in an image, etc.

- **Speech recognition** capability of a machine or a program to understand human speech like transforming human speech into a written format, or like smart home devices that take orders from humans and use speech recognition technology to understand their tasks.

- **Medical diagnosis** machine learning can be used to help diagnose disease like prediction and early diagnosis of breast cancer [82].

- **Predictive analytics** is the use of algorithms and machine learning techniques to determine future probabilities and predictions based on historical data. Predictive analytics can be used in marketing to find people whose data matches with ideal consumers and who are expected to purchase.

- **Traffic Prediction** Applications like Google Map help people to reach their destination easily showing them the correct path with the shortest route and predicts the traffic conditions. These applications uses real time location of the user and provide him with the better way to use and the average time of the journey according to the way he traveling.

- **Product recommendations** Various e-commerce and entertainment organizations, such as Amazon, Netflix, and others, employ machine learning to suggest product recommendations to users. Because of machine learning, whenever a person looks for a product on Amazon, he begins to receive advertisements for the same goods while browsing the internet on the same browser. Using multiple machine learning techniques, Google deduces the user's interests and recommends products based on those interests. Similarly, when we use Youtube, we receive recommendations for entertainment videos and content, which is also due to machine learning.

- **Email Spam and Malware Filtering** When we receive a new email, it is immediately categorized as important, normal, or spam. Machine learning is the technology that allows us to receive essential messages in our inbox with the important symbol and spam emails in our spam box. For email spam filtering and malware identification, machine learning algorithms such as Multi-Layer Perceptron, Decision Tree, and Naive Bayes classifier are utilized [8].

- **Online Fraud Detection** By detecting fraud transactions, machine learning makes our online transactions safer and more secure. When we conduct an online transaction, there are a number of ways for a fraudulent transaction to occur, including the use of phony accounts, fake ids, and the theft of funds in the middle of a transaction. To detect this, machine learning assists us by determining whether the transaction is genuine or fraudulent.

- **Automatic Language Translation** Nowadays in order to visit a new place, the language and the way to communicate is no longer an issue; machine learning can help expatriates and tourists to communicate by translating any text or speech into their original language and vice versa. This capability is provided by Google's GNMT (Google Neural Machine Translation), which is a Neural Machine Learning that automatically translates text into our native language [134].

2.4 What is Deep Learning?

Deep Learning is a class of machine learning that is commonly based on the model of neural networks. It enables the machine to imitate the actions of the human brain using artificial neural networks. Deep learning algorithms require much more training data than traditional machine learning algorithms. They are composed of layers that allow data to be passed between nodes(like neurons) in highly connected ways. The result is a non-linear transformation of the data that is increasingly abstract [76].

Deep learning currently represents the most advanced machine learning technique for a many high-level applications and tasks, especially for problems involving large structured training datasets [21]. It includes Recurrent Neural Networks(RNNs), Convolution Neural Networks(CNNs), and several other algorithms.

The relationship between artificial intelligence, natural language processing, machine learning, and deep learning is illustrated in Figure 2.2

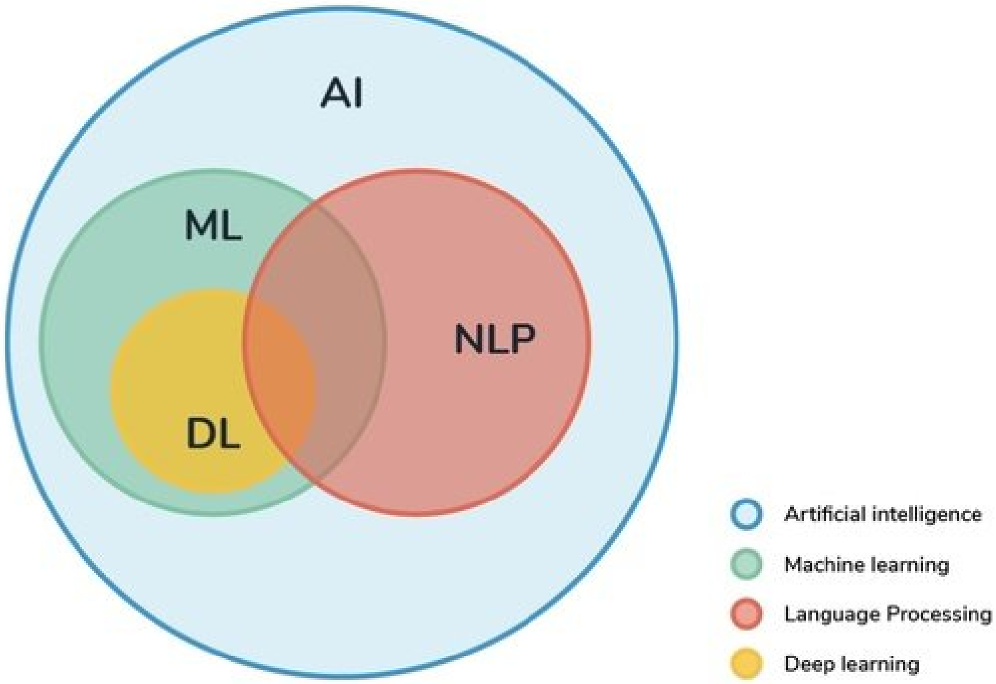


Figure 2.2: Relationship between AI, ML, DL, and NLP [4]

2.4.1 Neural Networks

Neural Networks, also known as artificial neural networks, are the heart of deep learning algorithms. Their name and structure are inspired from the human brain, mimicking the way biological neurons signal each other. Neural networks are comprised many layers, including an input layer, one or more hidden layers, and an output layer. Each layer consists of many nodes. These nodes, also called artificial neurons or units, are the basic computational component in a neural network. Neurons are linked one to another by weighted connections that form a network [83]. The architecture of a neural network is shown if Figure 2.3.

The calculation performed by the neural network involves values, called network weights or bias, which must be calculated and updated in such a way that the predictions are as true as possible. Data present in the input layer provide nodes with information in the forms of inputs. The inputs are multiplied with random weights, and a bias is added [84]. Finally, to decide which neuron to fire, nonlinear functions, also called as activation functions, are used. This is how the neural network perform the "learning" from the chosen examples.

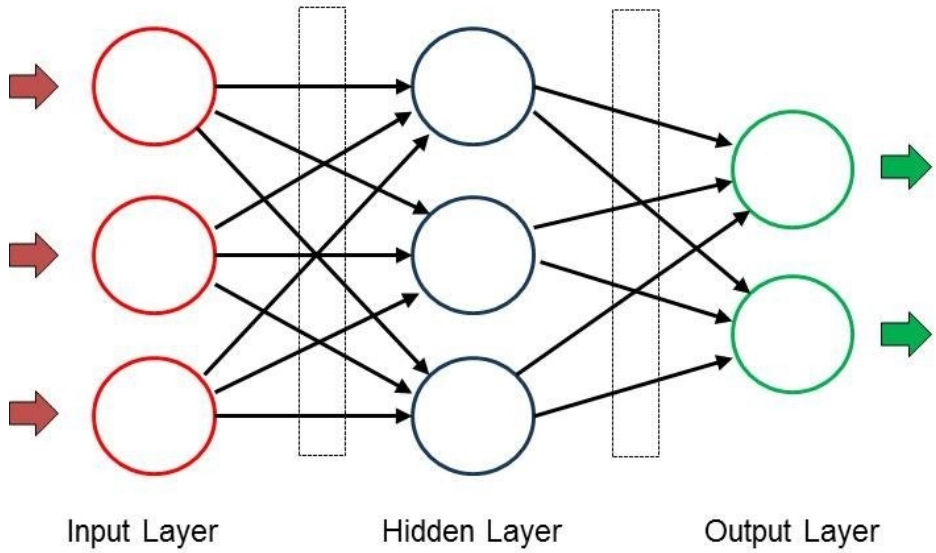


Figure 2.3: The architecture of a neural network [10]

Once the learning is complete, the network must be able not only to restore the examples of behavior learned, but above all to accurately predict unlearned behaviors [35]. Neural networks with multiple hidden layers are known as "deep" networks. They can learn more complex features and they are used in deep learning algorithms. Deep neural networks have forward propagation (or forward pass) and back propagation. The parameters of networks are updated and adjusted according to learning rate in order reach the minimized loss function.

2.4.2 Activation Functions

An activation function, also known as Transform function, is one of the most important parameters in a neural network. It adds non-linearity to the network and defines how the weighted sum of the input is transformed into an output from nodes in a layer of the network.

Mapping the input to the output is the core function of all types of activation function in all types of neural network. Non-linear activation layers are employed after all layers with weights (so-called learnable layers, such as fully connected layers and convolutional layers in CNN architecture). The most commonly used activation functions in deep neural networks are the following:

1. Sigmoid(Logistic Activation Function): The function takes any real value as input and outputs values in the range 0 to 1. The larger the input (more positive), the closer the output value will be to 1.0, whereas the smaller the input (more negative),

the closer the output will be to 0.0. It is especially used for models where we have to predict the probability as an output since probability of anything exists only between the range of 0 and 1. The sigmoid's mathematical representation is represented in Equation 2.1. Its curve looks like an S-shape and it is represented in Figure 2.4.

$$f(x) = \frac{1}{1 + e^{-x}} \tag{2.1}$$

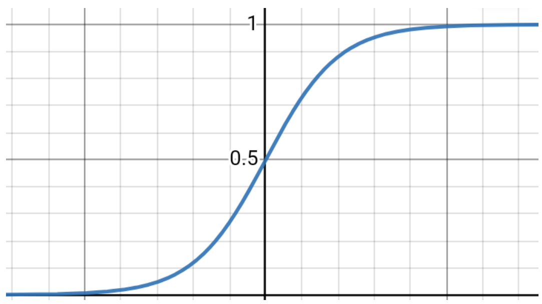


Figure 2.4: Sigmoid Activation Function

2. ReLU: Its range is from 0 to the infinity. The ReLU function, represented in figure 2.5, is half rectified (from bottom). It converts the whole values of the input to positive numbers. Equation 2.2 is the mathematical representation of ReLu function. F(x) takes zero when the x is less than zero and f(x) is equal to x when x is above or equal to zero. Lower computational load is the main benefit of ReLU over the other activation functions [7].

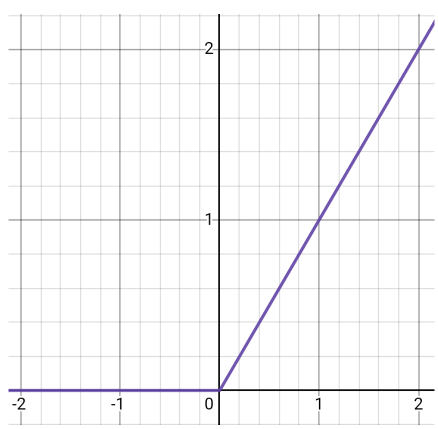


Figure 2.5: ReLu activation function

$$f(x) = max(0, x) \tag{2.2}$$

3. Tanh(Hyperbolic tangent Activation Function): It is similar to the sigmoid function(it is also S-shaped), as its input is real numbers, but the output is restricted to between -1 and 1. The larger the input (more positive), the closer the output value will be to 1.0, whereas the smaller the input (more negative), the closer the output will be to -1.0.
   The tanh function is mainly used classification between two classes [107]. Equation 2.3 is the mathematical representation of Tanh function, and its curve is represented in figure 2.6

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{2.3}$$

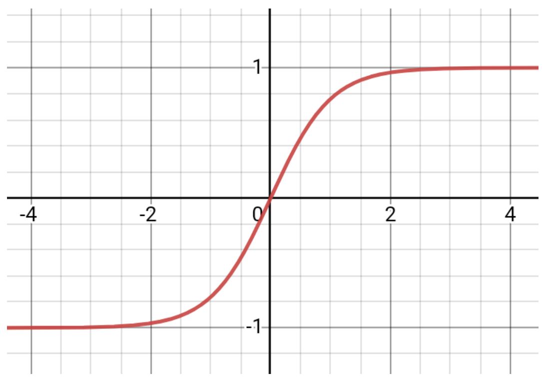


Figure 2.6: Hyperbolic tangent Activation Function

2.4.3 Backward Propagation

Backpropagation, short for backward propagation of errors, is a widely used algorithm in training feed-forward neural networks. It is the key to supervised learning of deep neural networks. When training a neural network, a loss function is calculated to estimate how far the network's predictions are from the true labels.

Backpropagation involves the calculation of the gradient proceeding backwards through the feed-forward network from the last layer through to the first. It allows us to calculate the gradient of the loss function with respect to each of the weights of the network. This enables every weight to be updated individually to gradually reduce the loss function over many training iterations [133]. The most common loss functions used are: Cross-Entropy, Euclidean Loss Function, and Hinge Loss Function.

2.5 Deep Learning Models

There are so many algorithms in the field of deep learning, we mention:

2.5.1 Recurrent neural networks

Recurrent neural networks (RNNs) are a class of neural networks that is most commonly used for sequence prediction problems. They have particularly gained popularity in the domain of natural language processing [51] like handwriting recognition [47] or speech recognition [46]. RNNs are widely used in the domain of deep learning because they can handle sequential data and they memorize previous inputs.
What differentiate an RNN from the other neural networks is their internal memory. RNN uses hidden layers to remember the sequence of words (data) and use the sequence pattern for the prediction. It uses an internal memory to remember the important features of the input they received. An ordinary feed-forward network only considers the current input, it can't remember anything about what happened in the past except its training. But in RNN, the information cycles through a loop. When it makes a decision, it considers the current input and the experience already learnt from the inputs received previously [33].This difference between a Feed-Forward Neural Network and a Recurrent Neural Network is illustrated in the figure 2.7.

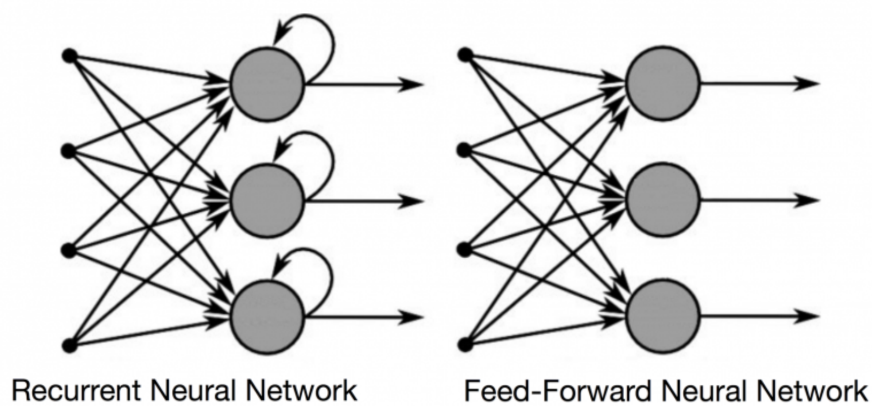


Figure 2.7: Recurrent Neural Network vs Forward Neural Network [34]

Feed-forward neural networks map one input to one output, but RNNs in addition can map one input to many outputs, many inputs to one output, and many inputs to many outputs. These types of RNN are illustrated in Figure 2.8.

However, two major difficulties have been found in training RNNs models:

- **Exploding gradients:** when the gradient becomes so large that the weights overflow the maximum limit [104]. Fortunately, this problem can be easily solved by truncating or squashing the gradients.

- **Vanishing Gradients:** Vanishing gradients occur when the values of a gradient are too small and the model stops learning or takes way too long as a result. The most

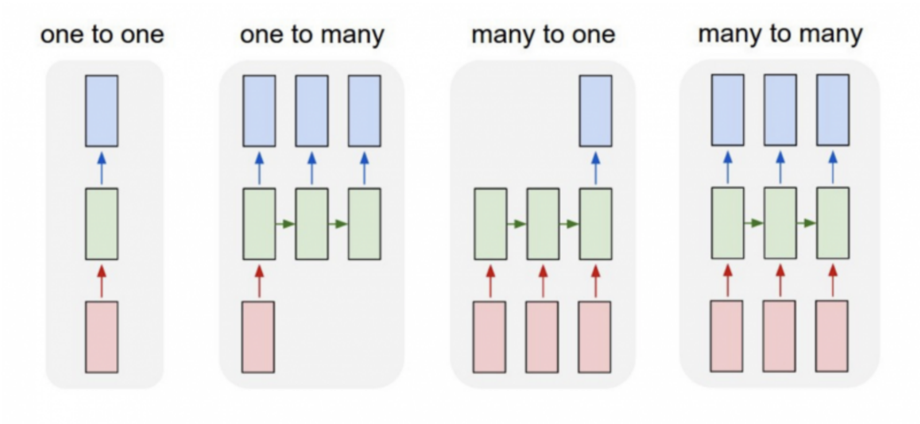


Figure 2.8: Recurrent Neural Network types [34]

effective solution to this problem is adding a gating mechanism to the RNN [104]. Two gated RNNs are found in the literature: Long short-term memory (LSTM) and Gated recurrent unit (GRU).

2.5.2 Long Short-Term Memory

RNNs have short memory.During the training process, the multiplications of several large or small derivatives may cause the gradients to exponentially explode or decay. With the entrance of new inputs, the network stops thinking about the initial ones; therefore, this sensitivity decays over time [7]. This issue can be handled using Long Short-Term Memory (LSTM)[52].

LSTM network is an alternative of RNN that extends the memory. It computes the input, output and forget gate to manage this memory. LSTM networks are well suited to learn from important experiences that have very long time lags in between. LSTM units can thus perform delicate tasks like propagating or keeping the flow of an important feature that came early compared to others in the input sequence worked over a long distance. The LSTM can read, write and delete information from its memory [7][104]. This memory can be seen as a gated cell, it means that the cell decides whether to open the gates or not (i.e., if it stores or delete the information) based on the importance it assigns to the information. The assigning of importance happens through weights, which are also learned by the algorithm. This simply means that it learns over time what information is important and what is not [33].

In an LSTM you have three gates: input, forget and output gate. These gates determine whether or not to let new input in (input gate), delete the information because it isn't important (forget gate), or let it impact the output at the current timestep (output gate). Figure 2.9 is an illustration of an LSTM with its three gates.

The gates in an LSTM are analog in the form of sigmoids, they range from zero to

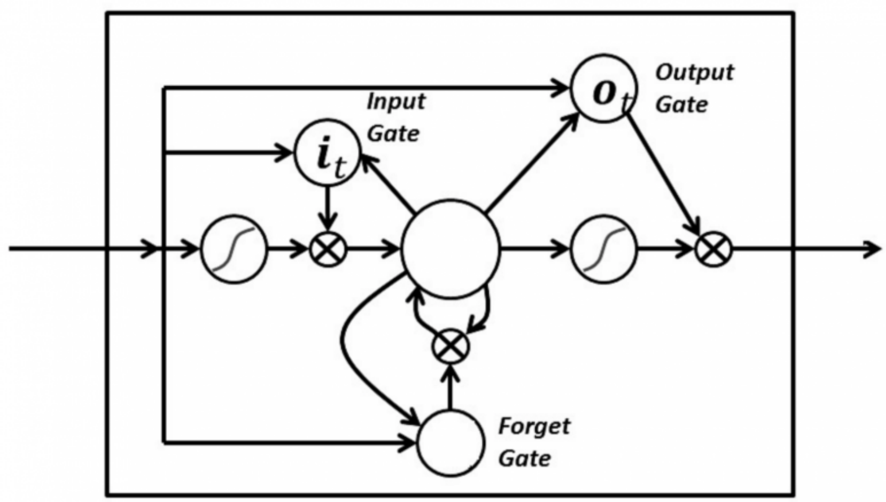


Figure 2.9: RNN with its three gates [34]

one. The fact that they are analog enables them to do backpropagation. The problematic issues of vanishing gradients is solved through LSTM because it keeps the gradients steep enough, which keeps the training relatively short and the accuracy high [33].

2.5.3 Gated Recurrent Unit

Like the LSTM, Gated Recurrent Unit(GRU) is also an improved version of standard RNNs to solve the vanishing gradient problem [122]. It consists of gating units that control the movement of data similar to the LSTM, but it performs this without making use of any extra memory cells.

GRU uses two important gates, called update gate and reset gate, to solve the vanishing gradient problem of the standard RNN. These gates modulate the movement of information passing into each hidden unit. The update gate is an important gate that is responsible for determining the amount of past information (from previous experience) that needs to be passed along the next state. The reset gate is used to determine how much of the past information to forget and neglect [41]. The GRU gates are illustrated in Figure 2.10

2.5.4 Convolutional Neural Network

One of the most popular and commonly employed deep neural networks is the Convolutional Neural Network (CNN) [86]. It takes its name from mathematical linear operation between matrices called convolution, which is the most significant characteristics of a CNN. CNNs are widely used in image classification and are the core of most computer vision systems today due to its ability to recognize patterns in images. However, we have recently

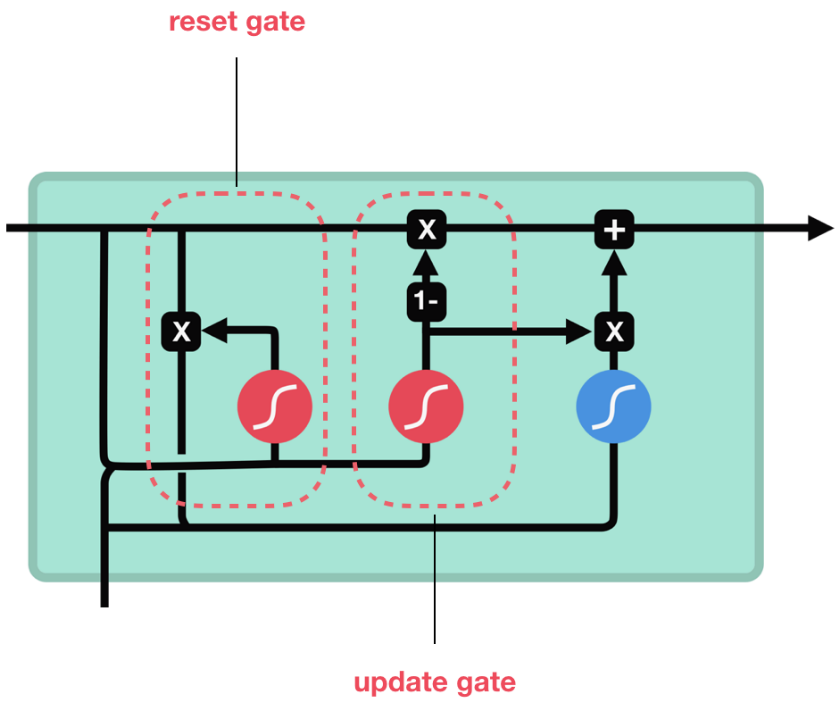


Figure 2.10: GRU cell and it's gates [91]

started also to apply CNNs to Sequence Processing problems and they achieved some interesting results [23].

CNN uses a special technique called Convolution. A convolution is the combination of two other functions to form a third function. It is applied to the input data with the help of a filter or kernel to build a feature map. By sliding the filter function on the input, we perform a convolution. Matrix multiplication is calculated at each point, and then the resulting is added to the feature map [29].

CNNs are comprised of three types of layers: convolutional layers, pooling layers, and fully-connected layers. Stacking these layers, a CNN architecture is formed. This architecture can be devided in two main parts:

- The first part in a CNN model is the feature extraction. The convolutional layer is used to detect the features. It is used to extract the spatial patterns in an input data by sliding a small kernel window over it. The network perform a multiple convolution operations on the input using different filters of a particular size (MxM) and generating various feature maps that give us information about the data entered. Then an activation function is applied to the convolution's output in order to present the non-linearity. Further shrinkage of the feature maps may appear since the size of input is larger than the size of the feature maps. In order to prevent this shrinkage, we use the padding.

Often, a convolutional layer is followed by a pooling layer. The main task of the pooling layer is decreasing the size of the convolved feature maps preserving their important characteristics. It creates smaller feature maps and maintains the majority of the dominant information (or features). Later, these feature maps are fed to other layers to learn several other features of the input. The sequence of convolution layer and pooling layer can be repeated many times. Depending on method used, there exist many types of pooling operations available for utilization in various pooling layers. The most familiar pooling methods are: Max Pooling, Min Pooling, and Average Pooling. An example of Max pooling is shown in Figure 2.11

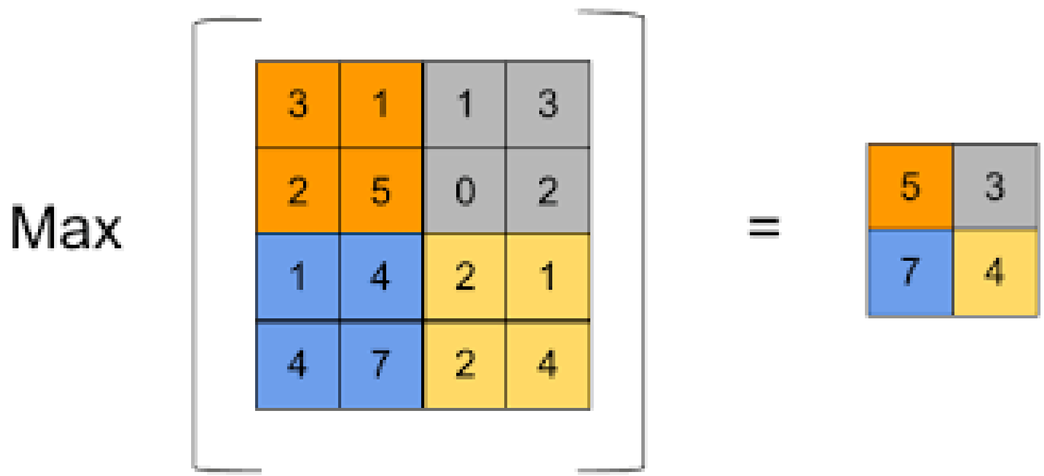


Figure 2.11: An example of Max Pooling [37]

- The second part in a CNN model is the classification part. First a flatten layer is presented to convert the 3D data resulted from the convolution layers into 1D data in order to be used in the fully connected layer. This layer is located at the end of each CNN model. It generates neurons that are connected to all neurons of the previous layer. It functions as a classifier and give the input data its classification predicted by the CNN model. The input of this layer is produced at the end of the last pooling layer, flattened, and fed to the FC layer in the form of a vector. The Convolutional neural network can contain more than one Fully connected layer. The output of the last fully-connected layer classifies the data and represents the final CNN output.

The CNN architecture is illustrated in Figure 2.12

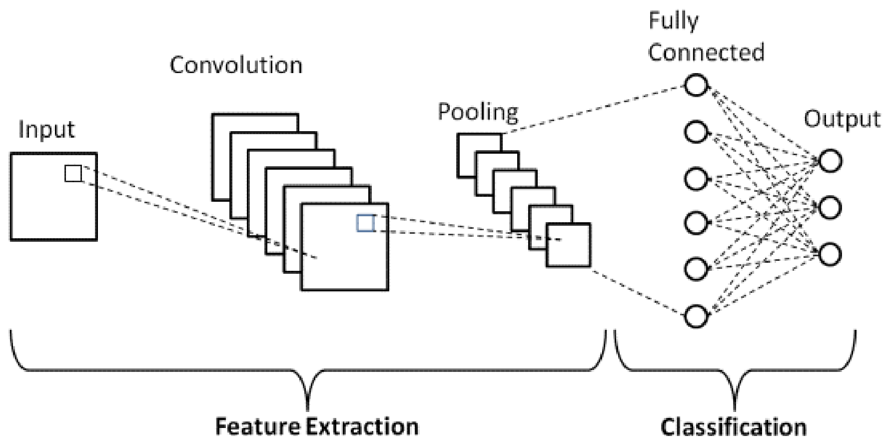


Figure 2.12: CNN architecture [57]

2.5.5 Generative Adversarial Networks

Generative Adversarial Networks (GANs) belong to the set of generative models. They have the ability to generate and produce new content. Given a training set, GANs learn to create new data instances with the same statistics as the training set. For example, a GAN trained on photographs can generate new photographs that have many realistic characteristics and seem at least superficially reliable to human observers [42].
GAN consists of two components: a generator that learns to generate fake data, and a discriminator that is able to tell how much an input is "realistic" by learning from that false information. The discriminator learns to differentiate between the generator's fake data and the real sample data. The generator produces fake data during the early training, and the discriminator quickly learns to recognize it as false. To update the model, the GAN transfers the results to the generator and discriminator [121]. The Figure 2.13 is a diagram illustrating how GANs work.

2.6 Ensemble Learning

Machine learning algorithms have their limitations and producing a model with high accuracy is challenging. Certain models do well in modeling one aspect of the data while others do better in modeling others. Ensemble learning aims to train multiple learners to solve the same problem.

Ensemble learning is also called commitee-based-learning or learning multiple classifier systems [140]. It refers to algorithms that combine multiple other models together in the prediction process in order to improve the stability and the predictive performance. Those models are referred to as base learners or base estimators. It overcomes some technical challenges of building a single model like high variance, low accuracy or features noise and bias. It will provide a composite prediction where the final accuracy is better than the

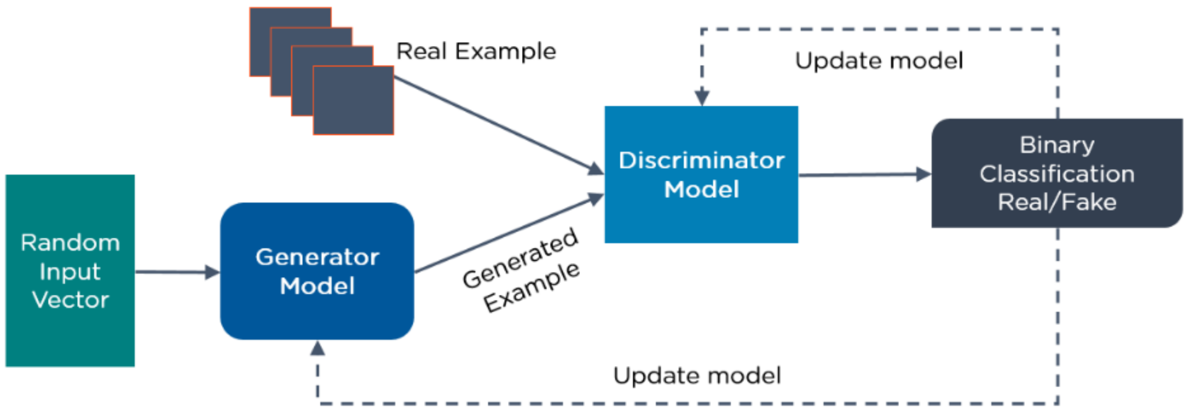


Figure 2.13: Generative Adversarial Network [120]

accuracy of other models.

En general, Ensemble learning is constructed in two steps: building the base learners and combining the results. To get good ensemble, base learners should be accurate enough (more than 50%) and diverse as possible [140]. The three main techniques of ensemble learning are bagging, stacking, and boosting.

2.6.1 Bagging Ensemble Learning

This typically involves using a single machine learning algorithm, almost always an unpruned decision tree, fitting many decision trees on different subsets of the same dataset and averaging the predictions. Each model learns the error produced by the previous model using a slightly different subset of the training dataset. Each subset has the same equal size and can be used to train models in parallel.

Bagging reduces variance and minimizes overfitting. One example of such a technique is the Random Forest algorithm. The Figure 2.14 shows the architecture of bagging ensemble.

2.6.2 Boosting Ensemble Learning

The principle property of boosting ensembles is the idea of correcting the errors of prediction. It trains base learners sequentially, each trying to correct its predecessor. The models are fit and added to the ensemble sequentially so that the second model tries to correct the predictions of the first model, the third corrects the second model, and so on [14]. The Figure 2.15 shows the architecture of boosting ensemble.

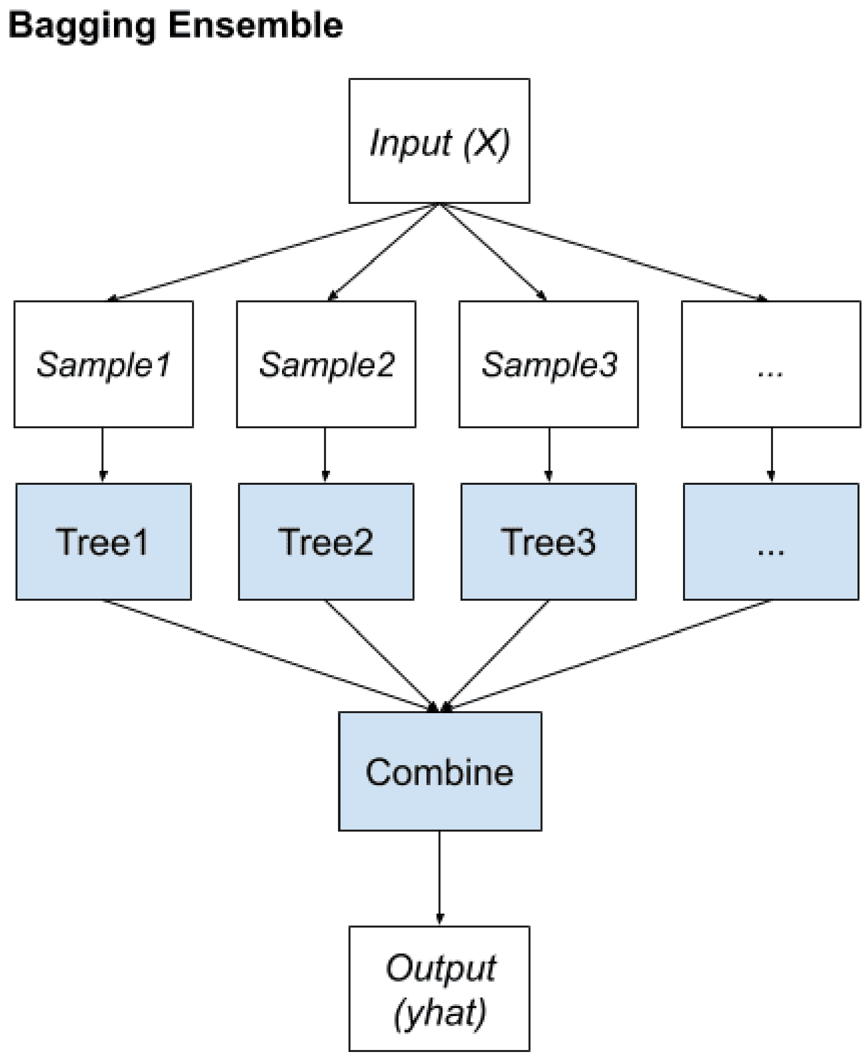


Figure 2.14: Bagging Ensemble [14]

2.6.3 Stacking Ensemble Learning

Stacking is a general process of learning how to combine the individual learners and create such a stronger model from all weak learners' predictions. Here, the individual learners are called the first-level learners, while the combiner is called the second-level learner, or meta-learner [140].

Stacking ensemble learning uses different machine learning algorithms for each ensemble member but all of them are trained over the same dataset. After that, a machine learning model is applied to learn how to best combine predictions and get the final prediction depending on the ensemble estimators. The Figure 2.16 shows the architecture of stacking

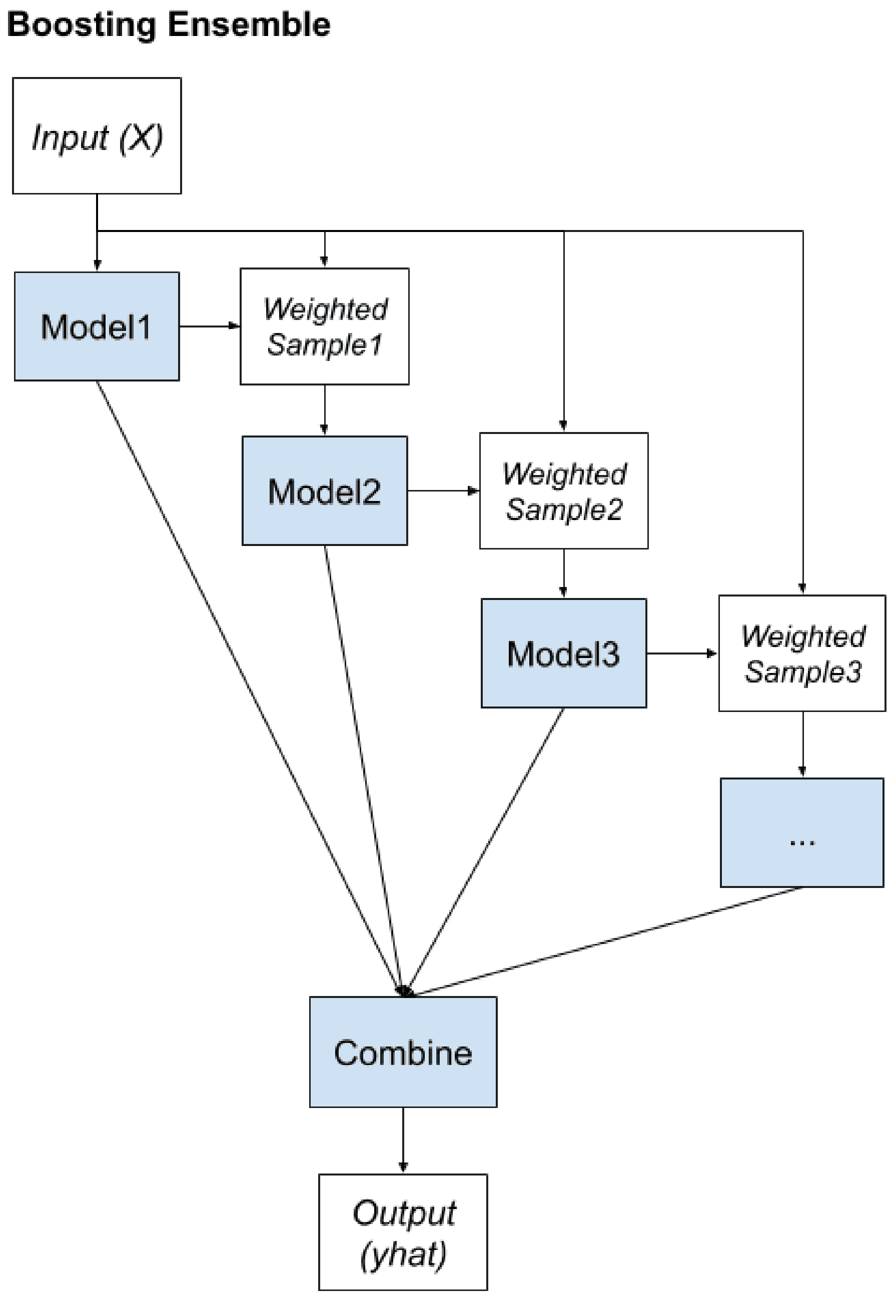


Figure 2.15: Boosting Ensemble [14]

ensemble.

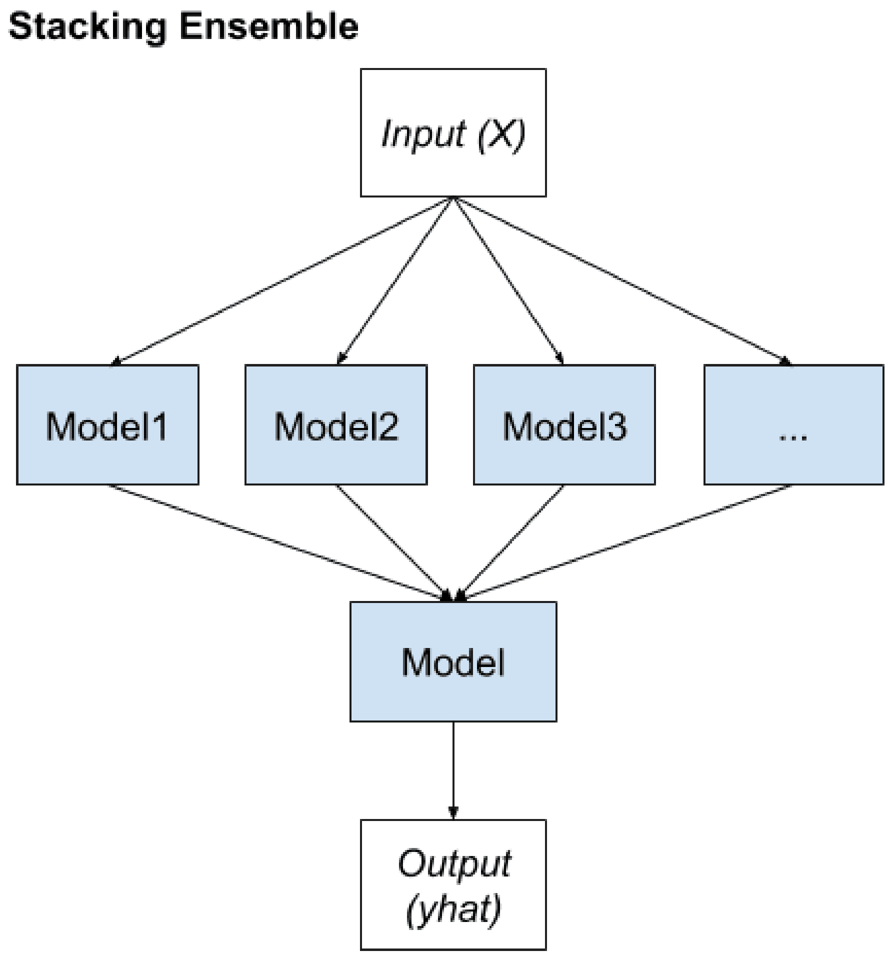


Figure 2.16: Stacking Ensemble [14]

2.7 Conclusion

In this chapter we have defined the concepts artificial intelligence, natural language processing, machine learning, deep learning, and ensemble learning. We also introduced a set of deep learning algorithms such as convolution neural networks(CNN), Long Short Term Memory(LSTM), Gated Recurrent Units(GRU), and Generative Adversarial Network(GAN). We will implement some of these algorithms for the application of our system along with the concept of ensemble learning that will help combine these algorithms together. In the next chapter, we will explain in detail the architecture and the implementation of our model that we have built for fake news detection.

Chapter 3

Conception, implementation and results

3.1 Introduction

A huge amount of fake news was spread in social media during the Covid-19 pandemic, and this has posed a serious problem for public health. Machine learning provides us with many advanced techniques that enable us to tackle this infodemic and reduce the spread of fake news in social media. In this chapter we will introduce our work that is a model for detecting fake news shared on social media related to the COVID-19 pandemic. In the first section, 3.2, we will define the objective and motivations behind this work. In the second section, 3.3, we will cite some related works and then in Section 3.4, we will explain the conception of our model. Next, we will describe the implementation of our model in Section 3.5. And finally, we will discuss, in Section 3.6, the results obtained and compare them with other works using the same dataset.

3.2 Objective and Motivations

The advent of the internet has been a game changer. With its ubiquity and easy accessibility, the internet provides an immediacy of information no other news medium can match. This has irrevocably accelerated the pace of the news, it has made it easy for anyone to publish content online and potentially reach large audiences in real time. Similarly, the way people consume the news has changed. People have easy and rapid accessibility to a huge amount of information available 24/7 and they are no longer restricted to any morning papers or evening news broadcasts. On the other hand, the internet led to a huge explosion in the amount of false information and fake news created and spread online, it has become difficult to tell whether stories are credible or not.

Fake news has become prevalent in the age of social media and have known a huge increase since the beginning of the COVID-19 pandemic. Commonly known, COVID-19 Corona virus, emerged in Wuhan, China, in December 2019. In the wake of this pandemic, we have witnessed a massive infodemic. This infodemic poses serious problems for public health. So, fighting this infodemic has become a significant challenge. Therefore, a strong need emerges to find a suitable model that can easily and efficiently identify and detect fake news related to Covid-19.

The main goal of this work is to create a model that would be able to detect fake news on social media in the context of COVID-19 and to distinguish them from the real news. We use in this work the Constraint@AAAI 2021 Covid-19 Fake news dataset that contains news related to Covid-19, collected from various online social networks [88], and labeled as fake or real. We generate an ensemble of three different deep learning-based models to construct our system. The deep learning models used in our ensemble are: convolutional neural network,long short-term memory and Gated Recurrent Unit. Our system is capable to evaluate news and make a decision whether they are fake or real.

3.3 Related works

The Contraint@AAAI 2021 Covid-19 Fake news detection dataset was released as part of a shared task at CONSTRAINT workshop collocated with AAAI21 and many works have been released for fake news detection using this dataset. Sunil Gundapu and Radhika Mamidi [48] used ensemble learning. They built an ensemble of three transformer models (BERT, ALBERT, and XLNET) to detect fake news. Apurva wani and al. [128] used many classification algorithms that are based on Convolutional Neural Networks (CNN), Long Short Term Memory (LSTM), and Bidirectional Encoder Representations from Transformers (BERT). Mohammad Faiyaz Khan and al. [105] constructed a stacking ensemble with eight different transformer-based pre-trained models with additional layers. Prathmesh Pathwar and Simran Gill [87] used an ensemble model RMDL that combines between CNN, RNN, and DNN (Dense Neural Network). Elena Shushkevich and John Cardiff [111] constructed an ensemble consisting of Bidirectional Long Short Term Memory(LSTM), Support Vector Machine(SVM), Logistic Regression(LR), Naive Bayes(NB), and a combination of Logistic Regression and Naive Bayes. Andrea Stevens Karnyoto and al. [62] used an approach based on transfer learning.

3.4 Conception

In this work we propose a model capable of detecting fake news from social media. First we organize our data into two sets, the first for training and the second for testing. Then we apply data pre-processing (noise removal, stop words removal, and stemming) to both sets.

Next, we build three deep learning models: a Convolutional Neural Network (CNN), a Long Short-Term Memory (LSTM), and a Gated Recurrent Units(GRU). After building the models, we train them over the clean data from the training dataset, and we combine them into an ensemble model. Then, the model is ready to make a decision and predict

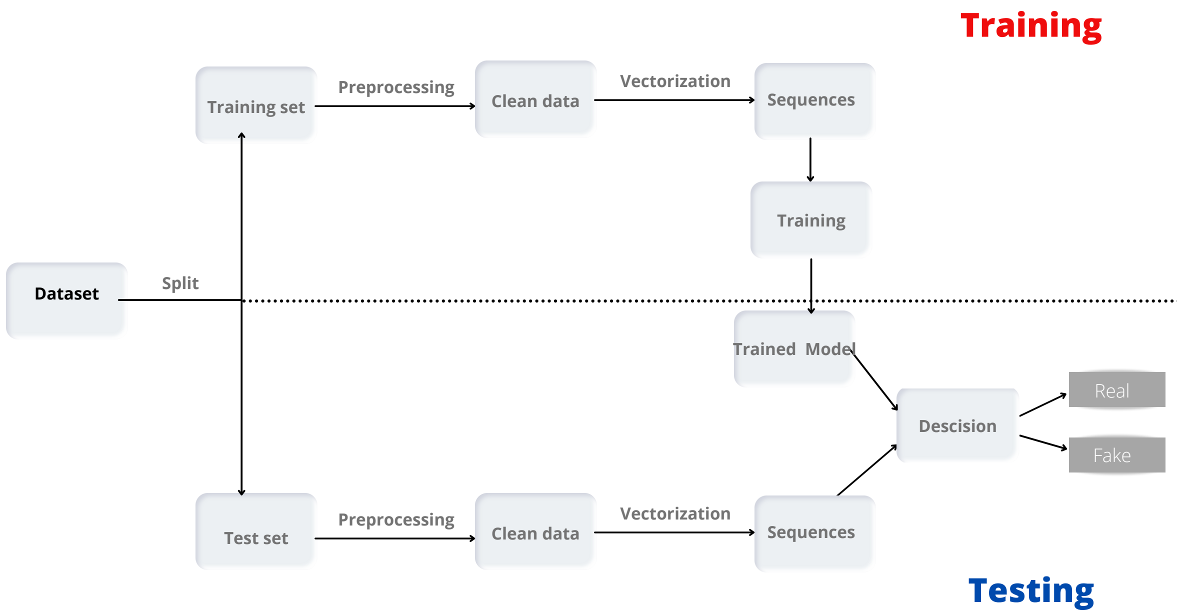


Figure 3.1: System architecture

the class of a news(fake or real). The architecture of our model is shown in Figure 3.1.

3.4.1 Data pre-processing

The first step is data preparation, our dataset contains thousands of claims written in the human language full of noise. This noise represents some distraction of the information presented in the claim and data processing consists on cleaning it in order to improve the quality of data.

Text data contains noise in various forms like links, punctuation, text in different case characters, etc. Before using this data for analysis or prediction, it must be cleaned. During this phase, we apply several techniques to clean the noise from the data and make it ready for training the model. This is an important step in natural language processing (NLP) tasks, it transforms the text into a more digestible form which makes machine learning algorithms able to perform better.

We apply the following steps for pre-processing our data:

- **Case conversion:** this step is very important because the same word with different case characters is seen as two different words for the machine (like: example and EXAMPLE). So, we convert the text into the same case to provide better understanding for the machine.

- **Remove Links:** in this step we will remove all the URL links and replace them by the string “url” in order to make them easier to identify.

- **Remove punctuation** in this step, all the punctuation is removed from the text. The library “string” of Python provides some pre-defined list of punctuation.
- **Reducing repeated characters:** there might be scenarios where characters are repeating more than necessary. We limit this repetition to two characters.
- **Remove stop Words:** this step consists on removing words that commonly occur in sentences and usually add less value to the overall meaning of the sentence. These words may be pronouns, conjunctions, prepositions, hyperlinks or other words that create unnecessary noise during the training of the model. Removing these words allows the model to focus on the words that convey the main focus of the sentence.
- **Stemming:** this process consists on removing the prefix or suffix from the word and reducing it to its root stem. This removes the multiple variation of words in different formats in order to eliminate the redundancy within the text for efficient and better learning.

3.4.2 Proposed model

In order to detect fake news on social media, we propose to use an efficient technique called ensemble learning. This technique consists in combining the outputs of various candidate models to deduce a final output. The ensemble consists of three deep learning models: Convolutional Neural Network, Long Short-Term Memory, and Gated Recurrent Unit. We separately train each one of the three deep learning models. The proposed ensemble model is shown in Figure 3.2.

After pre-processing our data and representing it as vectors, we build the models that can learn from it. The three models are compiled using ”binary_crossentropy” loss function and ”adam” optimizer.We describe below the varieties of deep learning models used in our approach:

CNN model

CNN is mostly used for image recognition, but they can have good performance on sentence classification as well [23]. The first layer in the Convolution Neural Network (CNN) is the embedding layer. This layer takes multiple parameters, the input dimension is used to configure the vocabulary size, the input length is used to configure the length of input sequences, and the output dimension is used to configure the size of the embedded words. Our embedding layer takes the input as a vector of 7400 word indices of length 20 and outputs a vector of 200 word indices. This layer converts the input texts n×m sequence matrix where n is the length of the input data and m is the length of the word embedding dimension. After the embedding layer, we have three 1D convolutions of kernel size 4. These convolutional layers are stacked up together with a max pooling layer following

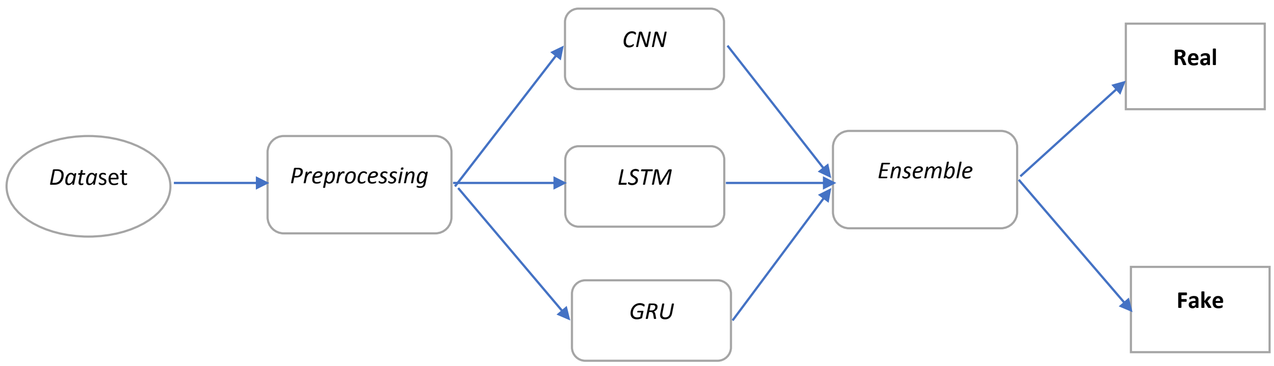


Figure 3.2: The ensemble model proposed

each one of them. The CNN layers extract learning local features (phrases) from the word embeddings to generate the output and max pooling can choose the most significant global feature and reduce the number of operations required for all the following layers in the network. The first convolutional layer receives input from the embedding layer. The kernel size of the three convolutional layers is 4 while the filter sizes for each layer are 128, 64, and 32 respectively. Next, we have a flatten layer as a function that converts the features taken from the pooling layer and maps them to a single column that is further passed to the fully connected layer. Finally, the model ends with two dense layers. A dense layer, so-called fully connected layer, functions as a linear operation in which every input is connected to every output by some weight. The first dense layer has 256 nodes and takes ReLu as an activation function, while the second functions as the output layer with 1 node and a sigmoid activation function.

The layered architecture of the CNN model is illustrated in the Figure 3.3.

LSTM model

LSTM is an extension of Recurrent Neural Network (RNN) that can solve long term dependency problem, and it is commonly used for sequence for sequence classification. First, we implemented an Embedding layer in which the input dimension, output dimension, and input length are 7400, 128, and 32 respectively. The output of this layer is passed as input

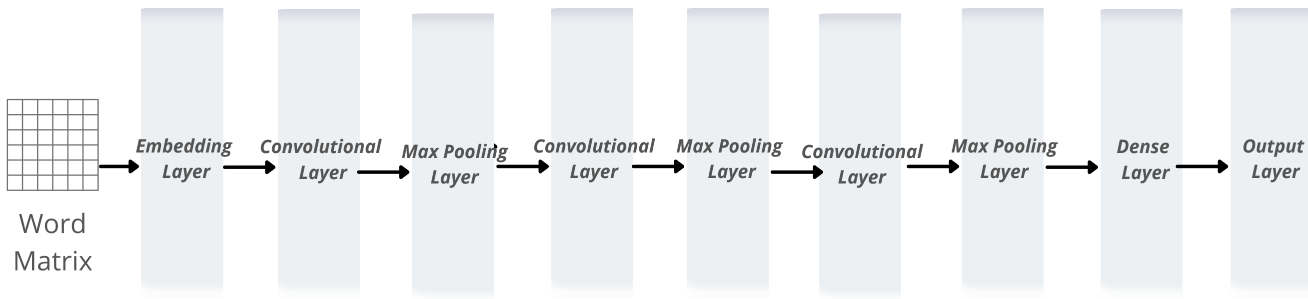


Figure 3.3: The architecture of the CNN model

to the LSTM layer. This layer can handle the nature of sequential data and catch the context of the input sentence by capturing long-term dependencies between word sequences. It takes the outputs of the embedding layer as input and it consists of 128 units. Then we have considered three dense layers in our LSTM network. The first dense layer has 128 nodes and ReLu as an activation function. The second dense layer has 64 nodes and ReLu activation function. The last dense layer functions as the output layer with 1 node that defines the value of the prediction. This layer is activated using sigmoid activation function and it takes binary cross-entropy as a loss function.

The layered architecture of the LSTM model is illustrated in Figure 3.4.

GRU model

Gated Recurrent Neural Network (RNN) is also an extension of Recurrent neural network (RNN) that eliminates the vanishing issue of gradient problem using update and reset gates[25]. The first layer in our GRU model is the embedding layer that has an input dimension of 7400, an output dimension of 200, and an input-length of 20. The second layer is GRU layer. The two main parts of the GRU are the update gate and the reset gate. A reset gate is used as a control mechanism to ignore the output of the hidden layer information or not. Likewise, an update gate is used to manage the degree of impact at

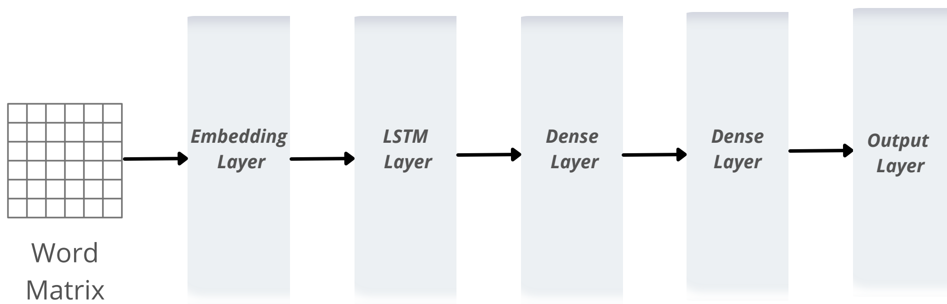


Figure 3.4: The architecture of the LSTM model

the previous time at the current hidden layer[62]. After the GRU layer, we add two dense layers, the first has 64 nodes and ReLu activation function while the second performs as an output layer with 1 node and sigmoid activation function.

The layered architecture of the GRU model is shown in Figure 3.5.

3.4.3 Training and validation

Training involves making a prediction based on the current state of the model, calculating how incorrect the prediction is by using the loss function, and updating the weights or parameters of the network according to the chosen optimizer to minimize the error and make the model predict better.

We train the constructed models over the pre-processed data. We apply the training using "fit" method. The batch size used in the three models is 32 which represents the number of samples from the training data that are used to compute the loss, and the weights are updated once based on this value. The number of epochs is 20 which is the number of times the the training will repeat over the entire dataset. At the end of each epoch, we use the validation set to evaluate how well the model is learning. We allocate 10% of the data for validation.

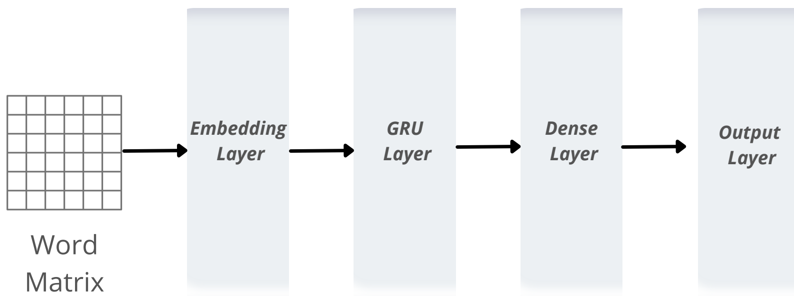


Figure 3.5: The architecture of the GRU model

3.4.4 Proposed ensemble model

Aiming for a better prediction, we combine all of the three deep learning-based models(CNN, LSTM, GRU) using a stacking ensemble. These models are called first-level learners, or estimators while the combiner is called the second-level learner, or meta-learner. Each one of the estimators predicts the probability of the news' class, either fake or real, then the predictions are combined by averaging them. If the average computed is higher than 0.5 the news is marked as real, else it's marked as fake news.

The algorithm used for the ensemble model is shown in Figure 3.6

3.5 Implementation

3.5.1 The dataset description

The Constraint@AAAI 2021 Covid-19 Fake news detection dataset, created by Parth and al.[88], is a dataset that consists of social media posts and news articles related to COVID-19 topic. It is written in English language and released as part of a shared task at CONSTRAINT workshop collocated with AAAI21. This challenge aims to help at detecting fake news about COVID-19 in social networks and stand against their wild

```
Algorithm : Ensemble Model Prediction
Input : news_claim
Output : fake, real
Variables : cnn_prediction, lstm_prediction, gru_prediction,ensemble_model_prediction
Begin
    cnn_prediction=CNN_model_prediction(news_claim)
    lstm_prediction=LSTM_model_prediction(news_claim)
    gru_prediction=GRU_model_prediction(news_claim)
    ensemble_model_prediction=( cnn_prediction+ lstm_prediction+ gru_prediction)/3
    if(ensemble_model_prediction > 0.5)
        return real
    return fake
End
```

Figure 3.6: The algorithm of the ensemble model proposed

spread.
The dataset contains 10700 of claims; each claim is labeled as real or fake. Fake claims are collected from various fact-checking websites like Politifact[1], News Checker [2], Boomlive[3], and from tools like Google fact-check-explorer[4] and IFCN chatbot. These claims were verified to be "not true". Real claims are collected from Twitter using verified twitter handles.

Table 3.1: Size and label distribution of data

	Real	**Fake**	**Total**
Train	3360	3060	6420
Validation	1120	1020	2140
Test	1120	1020	2140
Total	5600	5100	10700

[1] https://www.politifact.com/
[2] https://newschecker.in/
[3] https://www.boomlive.in/
[4] https://toolbox.google.com/factcheck/explorer

The dataset contains 10,700 media articles and posts collected from multiple platforms and categorized into 5600 real and 5100 fake claims. The dataset is split into 3 files: train, validation and test. The train data has 6420 claims, the number of fake claims is 3060 while the number of real claims is 3360. Each one of The validation data and the test data have 2140 claims, 1020 of them are fake and 1120 are real.

Table 3.1 shows the size and label distribution of data. We can observe that the number of fake claims and the number of real claims in each set are comparable which would give better insights about the performance of our models of detection. The table 3.2 shows some examples of fake and real claims from the dataset.

Table 3.2: Examples from dataset

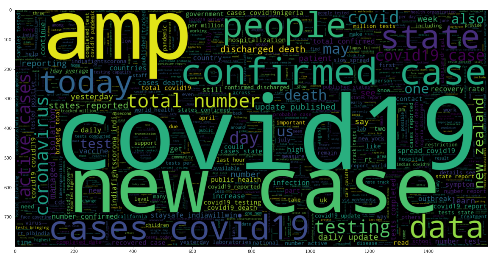
Tweet	**Label**
Covid Act Now found "on average each person in Illinois with COVID-19 is infecting 1.11 other people. Data shows that the infection growth rate has declined over time this factors in the stay-at-home order and other restrictions put in place." https://t.co/hhigDd24fE	real
Italy has surrendered to the coronavirus pandemic as all the measures to control COVID-19 have been exhausted.	fake
President Trump Says Pandas Are Responsible For Covid-19 https://t.co/mgan5G2oY8 #donaldtrump #china #coronavirus #zoo #pandas	fake
There are currently 4236. people in managed isolation and quarantine. Our current effective capacity is 7103. This gives us an excess capacity of 2867. Over the next week we are projecting 2789 arrivals and 1267 departures from our facilities.	real

After performing the pre-processing steps explained in Section 3.4.1, additional statistics of the dataset are gathered and presented in Table 3.3.

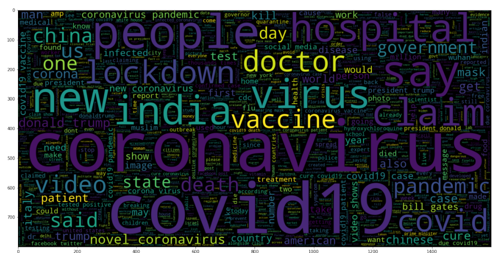
We highlight in Figure 3.7, using the word cloud, the most popular and occurring

Table 3.3: Statistics of the dataset

Feature	Train	Test
Total number of words	149276	38078
Total number of unique words	16551	7400
Maximum number of words in a document	902	982
Average number of words in a document	17.70	18.10



(a) Word cloud of real news



(b) Word cloud of fake news

Figure 3.7: Three simple graphs

words of the real and fake data after removing the stop words. The larger and bolder the word appears in the word cloud, the more frequent it's mentioned within the dataset and the more important it is. In Figure 3.7a, we can see words in real labeled data which don't often occur in Figure 3.7b like "discharged", "confirmed", "testing", and "active cases". While in Figure 3.7b, we can find the unique words frequently appearing in the fake articles but not frequently appearing in the true labeled articles like "government", "vaccine", "president donald", "lockdown", and "cure". These frequent textual words can give important information to differentiate the true data from the fake one.

3.5.2 Deep learning models

The three models are implemented using Keras library. The CNN model consists of 9 layers: an embedding layer, a convolutional layer with 128 filter size, a max pooling layer, a second convolutional layer with 64 filter size, a max pooling layer, another convolutional layer with 32 filter size, a max pooling layer, and 2 dense layers. Its implementation is shown in Figure 3.8

The LSTM model consists of 5 layers: an embedding layer, an LSTM layer with 128 units, and 3 dense layers with 128, 64, and 1 node in each layer respectively. Its implementation is shown in Figure 3.9

The GRU model consists of 4 layers: an embedding layer, a GRU layer, and 2 dense layers with 64 nodes and 1 node respectively. Its implementation is shown in Figure 3.10

```
embedding_vecor_length = 32
model_CNN = Sequential()
model_CNN.add(Embedding(total_words, 200, input_length=20))
model_CNN.add(Conv1D(filters=128, kernel_size=4, padding='same', activation='relu'))
model_CNN.add(MaxPooling1D(pool_size=2))
model_CNN.add(Conv1D(filters=64, kernel_size=4, padding='same', activation='relu'))
model_CNN.add(MaxPooling1D(pool_size=2))
model_CNN.add(Conv1D(filters=32, kernel_size=4, padding='same', activation='relu'))
model_CNN.add(MaxPooling1D(pool_size=2))
model_CNN.add(Flatten())
model_CNN.add(Dense(256, activation='relu'))
model_CNN.add(Dense(1, activation='sigmoid'))
model_CNN.compile(loss='binary_crossentropy', optimizer='adam', metrics=['accuracy'])
model_CNN.summary()
```

Figure 3.8: The implementation of the CNN model

```
modelLSTM = Sequential()
modelLSTM.add(Embedding(total_words,  output_dim=128,mask_zero=True))
modelLSTM.add(LSTM(128))
modelLSTM.add(Dense(128, activation='relu'))
modelLSTM.add(Dense(64, activation='relu'))
modelLSTM.add(Dense(1))


modelLSTM.compile(loss='binary_crossentropy', optimizer='adam', metrics=['accuracy'])
modelLSTM.summary()
modelLSTM.fit(padded_train, y_train, batch_size = 32, validation_split = 0.1, epochs = 20)
```

Figure 3.9: The implementation of the LSTM model

```
modelGRU = Sequential()
modelGRU.add(Embedding(total_words,  output_dim=128,mask_zero=True))
modelGRU.add(tf.keras.layers.GRU(units=32))
modelGRU.add(Dense(name='fc', units=32, activation='relu'))
modelGRU.add(Dense(name='classifier', units=1, activation='sigmoid'))
modelGRU.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])
modelGRU.summary()
```

Figure 3.10: The implementation of the GRU model

3.5.3 Software tools

Because of metrial limitations, the experiments were conducted using a Google Colab RAM 12 GB. For implementing our model we used Python 3 programming language, Tensorflow, Keras, and NLTK library.

Google Colab

Google Colab or Colaboratory is a free cloud service, offered by Google. It is based on Jupyter Notebook and intended for training and research in automatic learning. This platform makes it possible to train machine learning models directly in the cloud without needing to install anything but a browser [20]. As of October 13, 2018, Google Colab provides a single 12GB NVIDIA Tesla K80 GPU that can be used up to 12 hours continuously. Recently, Colab also started offering free TPU [43].

Keras

Keras is a deep learning API written in Python, running on top of the machine learning platform TensorFlow. Its primary author and maintainer is François Chollet, a Google engineer [64]. Keras is an open-source software library that was designed to enable fast experimentation with deep neural networks. It focuses on being simple, modular, extensible, and powerful. It is used by organizations and companies including NASA, YouTube, or Waymo [2].

Python

Python[117], created by Guido van Rossum and first published in 1991, is a high-level interpreted programming language. It is dynamically-typed and supports multiple programming paradigms including structured (particularly procedural), object-oriented and functional programming. It is one of the most popular programming languages and its design philosophy emphasizes code readability with the use of significant indentation [96]. Python is used for many applications such as software development, data analysis, or infrastructure management.

Natural Language Toolkit

Natural Language Toolkit (NLTK) is a free, open source, community-driven project. It is a Python package that can be used in Natural Language Processing by making natural human language usable by computer programs. NLTK provides a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning, wrappers for industrial-strength NLP libraries, and an active discussion forum. NLTK is available for Windows, Mac OS X, and Linux [32].

3.6 Results and Discussion

In order to evaluate our model, we test it over the test set provided in the Constraint@AAAI 2021 Covid-19 Fake news detection dataset. First we will define the metrics used to evaluate our model.

3.6.1 Evaluation

We evaluate the results using the following metrics:

Confusion Matrix

A Confusion Matrix, also known as an error matrix, is a tool for measuring the overall performance of a supervised Machine Learning model by checking in particular how often its predictions are accurate compared to reality in classification problems. The confusion matrix provides more insight into not only the performance of a predictive model, but also which classes are being predicted correctly, which incorrectly, and what type of errors are being made [16]. The simplest confusion matrix is for a binary classification problem, with negative(class 0) and positive(class 1) classes [16]. It is represented in Table 3.4.

Table 3.4: Confusion matrix

	Positive prediction	**Negative prediction**
Positive Class	True Positive (TP)	False Negative (FN)
Negative Class	False Positive (FP)	True Negative (TN)

Accuracy score

Accuracy, used in classification problems, is a metric used to tell the percentage of accurate predictions. We calculate it by dividing the number of correct predictions by the total number of predictions that the model has made. Equation 3.1 shows how accuracy is calculated.

$$Accuracy = \frac{Number of correct predictions}{Total number of predictions} \tag{3.1}$$

Precision

Precision quantifies the number of positive class predictions that actually belong to the positive class[16]. It can be calculated as in Equation 3.2.

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives} \tag{3.2}$$

Recall

Recall quantifies the number of positive class predictions made out of all positive examples in the dataset [16]. It can be calculated as in Equation 3.3.

$$Recall = \frac{TruePositives}{TruePositives + FalseNegatives} \tag{3.3}$$

F1 score

F-Measure provides a single score that balances both the concerns of precision and recall in one number [16]. It can be calculated as in Equation 3.4.

$$F1score = 2 * \frac{1}{\frac{1}{Recall} + \frac{1}{Precision}} \tag{3.4}$$

3.6.2 Results

In this research, we employed three base learning (LSTM, GRU, CNN) and an ensemble that combines between them. The models are tested on the testing set provided in the Constraint@AAAI 2021 Covid-19 Fake news detection dataset. We evaluated their performance using four measurements which are accuracy, precision, recall, and the F1-score. The results are reported in Table 3.5.

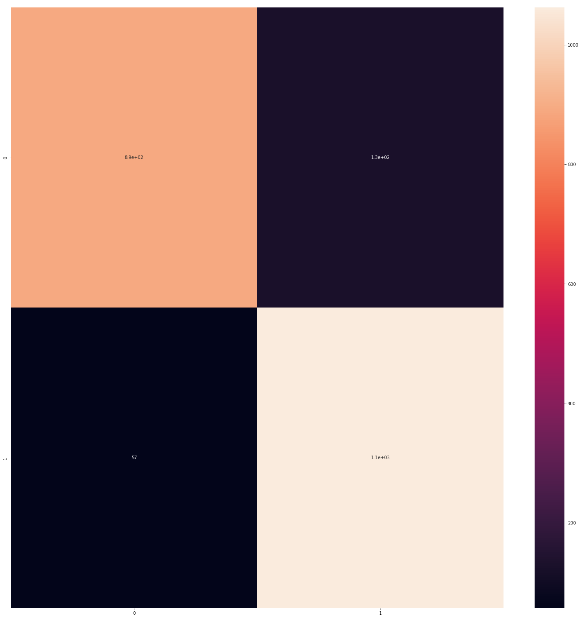
According to the results, the ensemble scored the greatest accuracy of 93.97%. Meanwhile, LSTM achieved 93.22%, GRU achieved 93.31%, and CNN achieved 91.44% accuracy rate. The results clearly demonstrate that models CNN, LSTM, and GRU are very good at classifying fake news related to Covid-19 subject, but the ensemble that combined these three models outperformed all of them with an accuracy of 93.97%. Note that all the results obtained are quite close, within 3% of each other.

For CNN, the precision, recall, and F1-score reached 89.40%, 94.91%, and 92.07% respectively. LSTM improved its precision, recall, and F1-score by 93.22%, 92.13%, and 95.17% respectively. GRU achieved 92.44% precision, 95.00% recall, and 93.70% f1-score, while the proposed ensemble model achieved 92.75% precision, 95.98% recall and 94.33% f1-score. Consequently, we find that the CNN-LSTM-GRU ensemble is the most efficient model for detecting fake news related to COVID-19 pandemic.

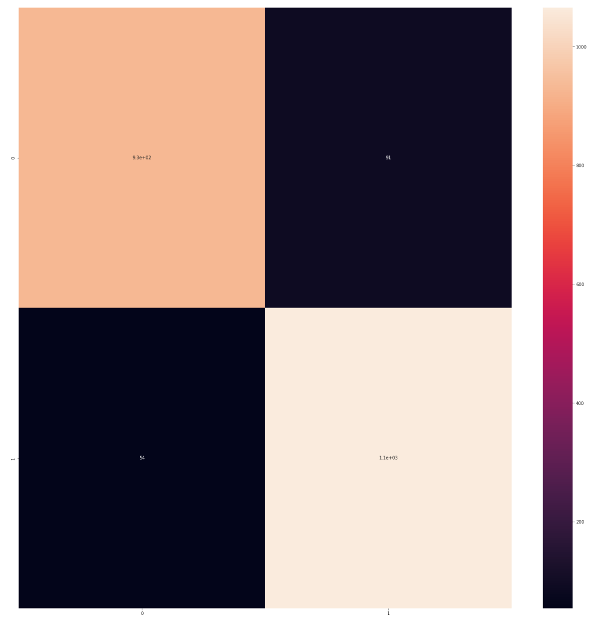
Table 3.5: Results

	Accuracy	**Precision**	**Recall**	**F1 score**
CNN	91.44	89.40	94.91	92.07
LSTM	93.22	92.13	95.17	93.63
GRU	93.31	92.44	95.00	93.70
Proposed ensemble	**93.97**	**92.75**	**95.98**	**94.33**

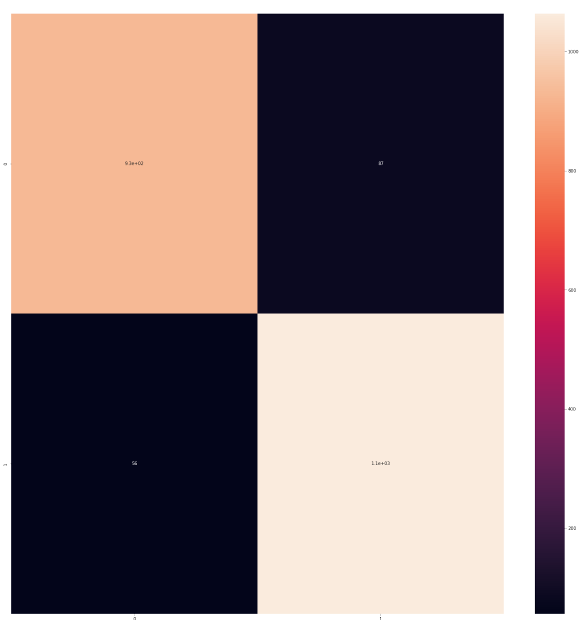
The confusion matrices, in Figure 3.11, show the overall performance of each base learner and the ensemble model.



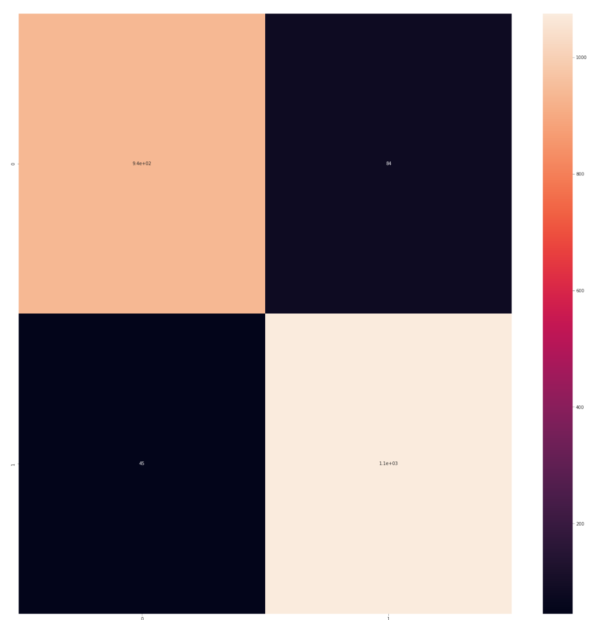
(a) Confusion matrix of CNN model



(b) Confusion matrix of LSTM model



(c) Confusion matrix of GRU model



(d) Confusion matrix of the ensemble model

Figure 3.11: Confusion matrices of the base learners and the ensemble model

3.6.3 Comparison with other related works

In this section, we compare the results obtained from our proposed ensemble model using Constraint@AAAI 2021 Covid-19 Fake news detection dataset with the other works using the same dataset.

Table 3.6: Comparison of results with other works

Work	Accuracy	F1 score
Apurva wani and al. [128]	98.41	/
Mohammad Faiyaz Khan and al. [105]	98.40	/
Sunil Gundapu and Radhika Mamidi [48]	98.55	/
Andrea Stevens Karnyoto and al. [62]	91.96	/
Prathmesh Pathwar and Simran Gill [87]	93.92	/
Elena Shushkevich and John Cardiff [111]	/	94.00
Our proposed ensemble model	93.97	94.33

Sunil Gundapu and Radhika Mamidi [48] achieved 98.55% which is the highest accuracy in this comparison. Mohammad Faiyaz Khan and al. [105] and Apurva wani and al. [128] come next with an accuaracy of 98.40% and 98.41% respectively. In [111], Elena Shushkevich and John Cardiff's model achieved an F1-score of 94% which is very close to the score achieved by our proposed model. Andrea Stevens Karnyoto and al. in [62] and Prathmesh Pathwar and Simran Gill in [87] achieved an accuracy of 91.96%¨and 93.92% respectively. Our proposed model outperformed both of them with an accuracy of 93.97%.

3.7 Conclusion

In this chapter, we presented, using the Constraint@AAAI 2021 Covid-19 Fake news dataset, an ensemble of deep learning-based models to combat the global infodemic related to Covid-19 pandemic. We cited some related works, we explained in detail the conception of our model, we described its implementation in detail, we discussed the results obtained from testing experiments and we compared them with other works that used the same dataset. The results we obtained indicate that, in our case, the ensemble of deep learning models outperforms these models when they are used alone. The ensemble model composed of CNN, LSTM, and GRU scored the highest accuracy of 93.97% which makes it very efficient for classifying fake news about COVID-19.

General Conclusion

Due to the rapid dissemination of fake news on social media, it is becoming very difficult to determinate the credibility and the truthfulness of the information presented to us. After the emergence of the Covid-19 virus in December 2019, fake news has known a huge increase that led to the creation of a massive infodemic. This infodemic has posed serious problems especially for public health [92]. It has caused a lot of panic, psychological issues, and inappropriate protective measures among people. In addition, this infodemic has cased a lack of trust in health institutions and hospitals [92]. Therefore, fighting this infodemic has become a significant challenge for researchers.

Detecting fake news has become a major challenge in today's world. In this our work, we proposed a model for detecting fake news shared on social media pertaining to the COVID-19 pandemic. Our approach is based on assembling three deep learning models: Convolutional Neural Network, Long Short-Term Memory, Gated Recurrent Unit.

The first chapter was divided into two sections.The first section was a definition of false information, its different types, actors behind it, and their different motives. In the second section, fake news was defined in detail with its types, methods of detection, and its impacts in the health stage.

The second chapter is an introduction to the concepts artificial intelligence, natural language processing, machine learning, deep learning, and ensemble learning. We also defined some commonly used deep learning algorithms like Convolution Neural Network(CNN), Long Short Term Memory(LSTM), Gated Recurrent Unit(GRU), and Generative Adversarial Network(GAN).

In the third chapter, first we defined the objective and motivations behind this work. Second, we cited some related works and then we presented the conception of our work. Next, we explained its implementation with a detailed description of the dataset used. Finally, we discussed the results obtained and compared them with other works using the same dataset. The ensemble model used in this work performed very well and achieved a high accuracy of 93.97%.

In the future we aim to use a larger dataset having larger set of vocabulary. We also intend to investigate the performance of new ensemble learning methods using different deep learning models for a better fake news detecting system. We are also planning to take in consideration other features that would improve the quality of detection like user-based features.

References

[1] Hervé Abdi, Dominique Valentin, and Betty Edelman. *Neural networks*. 124. Sage, 1999.

[2] *About Keras*. URL: https://keras.io/about/. (accessed: 11.05.2022).

[3] Wasim Ahmed et al. "COVID-19 and the 5G conspiracy theory: social network analysis of Twitter data". In: *Journal of medical internet research* 22.5 (2020), e19458.

[4] *AI, Machine Learning (ML) and Natural Language Processing (NLP)*. URL: https://athenatech.tech/f/ai-machine-learning-ml-and-natural-language-processing-nlp. (accessed: 08.06.2022).

[5] Sarah A Alkhodair et al. "Detecting breaking news rumors of emerging topics in social media". In: *Information Processing & Management* 57.2 (2020), p. 102018.

[6] Hunt Allcott and Matthew Gentzkow. "Social media and fake news in the 2016 election". In: *Journal of economic perspectives* 31.2 (2017), pp. 211–36.

[7] Laith Alzubaidi et al. "Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions". In: *Journal of big Data* 8.1 (2021), pp. 1–74. DOI: 10.1186/s40537-021-00444-8. URL: https://doi.org/10.1186/s40537-021-00444-8.

[8] *Applications of Machine learning*. URL: https://www.javatpoint.com/applications-of-machine-learning. (accessed: 06.06.2022).

[9] *Artificial Intelligence*. URL: https://fr.wikipedia.org/wiki/Artificial_Intelligence.

[10] *Artificial Intelligence for Beginners (2): Deep Learning*. URL: https://www.elektormagazine.com/articles/artificial-intelligence-for-beginners-2. (accessed: 08.06.2022).

[11] Barlamane.com. *En Algérie, l'appareil sécuritaire utilise l'affaire Djamel Ben Ismaïl à des fins de propagande*. URL: https://www.barlamane.com/fr/en-algerie-lappareil-securitaire-utilise-laffaire-djamel-ben-ismail-a-des-fins-de-propagande/. (accessed: 08.03.2022).

[12] Barlamane.com. *Exposing Russia's Effort to Sow Discord Online: The Internet Research Agency and Advertisements*. URL: https://intelligence.house.gov/social-media-content/. (accessed: 13.03.2022).

[13] Barlamane.com. *Useful idiots*. URL: https://rationalwiki.org/wiki/Useful_idiot. (accessed: 08.03.2022).

[14] Jason Brownlee. *A Gentle Introduction to Ensemble Learning Algorithms*. URL: https://machinelearningmastery.com/tour-of-ensemble-learning-algorithms/. (accessed: 08.05.2022).

[15] Jason Brownlee. *Deep learning for natural language processing: develop deep learning models for your natural language problems*. Machine Learning Mastery, 2017. ISBN: 9780198520115.

[16] Jason Brownlee. *How to Calculate Precision, Recall, and F-Measure for Imbalanced Classification*. URL: https://machinelearningmastery.com/precision-recall-and-f-measure-for-imbalanced-classification/#:~:text=Precision%5C%20quantifies%5C%20the%5C%20number%5C%20of,and%5C%20recall%5C%20in%5C%20one%5C%20number. (accessed: 05.05.2022).

[17] Erica Brownstein and Robert Klein. "Blogs". In: *Journal of College Science Teaching* 35.6 (2006), pp. 18–22.

[18] Zhan Bu, Zhengyou Xia, and Jiandong Wang. "A sock puppet detection algorithm on virtual spaces". In: *Knowledge-Based Systems* 37 (Jan. 2013), pp. 366–77. DOI: 10.1016/j.knosys.2012.08.016.

[19] Erin E Buckels, Paul D Trapnell, and Delroy L Paulhus. "Trolls just want to have fun". In: *Personality and individual Differences* 67 (2014), pp. 97–102.

[20] *C'est quoi Google Colab ?* URL: https://ledatascientist.com/google-colab-le-guide-ultime/. (accessed: 11.05.2022).

[21] Guillaume Chassagnon et al. "Deep learning: definition and perspectives for thoracic imaging". In: *European radiology* 30.4 (2020), pp. 2021–2030. DOI: 10.1007/s00330-019-06564-3. URL: https://doi.org/10.1007/s00330-019-06564-3.

[22] Cheng Chen et al. "Battling the internet water army: Detection of hidden paid posters". In: *2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2013)*. IEEE. 2013, pp. 116–120.

[23] Yahui Chen. "Convolutional neural network for sentence classification". MA thesis. University of Waterloo, 2015.

[24] Yimin Chen, Niall J Conroy, and Victoria L Rubin. "Misleading online content: recognizing clickbait as" false news"". In: *Proceedings of the 2015 ACM on workshop on multimodal deception detection*. 2015, pp. 15–19.

[25] Junyoung Chung et al. "Empirical evaluation of gated recurrent neural networks on sequence modeling". In: *arXiv preprint arXiv:1412.3555* (2014).

[26] Bryn Alexander Coles and Melanie West. "Trolling the trolls: Online forum users constructions of the nature and properties of trolling". In: *Computers in Human Behavior* 60 (2016), pp. 233–244.

[27] Lynne Connelly. “Logistic regression”. In: *Medsurg Nursing* 29.5 (2020), pp. 353–354.

[28] Jessie Daniels. “Cloaked websites: propaganda, cyber-racism and epistemology in the digital era”. In: *New Media & Society* 11.5 (2009), pp. 659–683.

[29] Merlin Susan David and Shini Renjith. “Comparison of word embeddings in text classification based on RNN and CNN”. In: *IOP Conference Series: Materials Science and Engineering*. Vol. 1187. 1. IOP Publishing. 2021, p. 012029.

[30] *Decision tree*. URL: https://en.wikipedia.org/wiki/Decision_tree. (accessed: 08.06.2022).

[31] Arwinder Dhillon and Ashima Singh. “Machine learning in healthcare data analysis: a survey”. In: *Journal of Biology and Today's World* 8.6 (2019), pp. 1–10.

[32] *Documentation*. URL: https://www.nltk.org/. (accessed: 06.06.2022).

[33] Niklas Donges. *A Guide to RNN: Understanding Recurrent Neural Networks and LSTM Networks*. 2021. URL: https://builtin.com/data-science/recurrent-neural-networks-and-lstm. (accessed: 28.02.2022).

[34] Niklas Donges. *A Guide to RNN: Understanding Recurrent Neural Networks and LSTM Networks*. URL: https://builtin.com/data-science/recurrent-neural-networks-and-lstm. (accessed: 08.06.2022).

[35] Gérard DREYFUS. *RÉSEAUX DE NEURONES*. URL: https://www.universalis.fr/encyclopedie/reseaux-de-neurones-formels/. (accessed: 28.04.2022).

[36] John Esterbrook. *U.S. Propaganda Push In Iraq*. URL: https://www.cbsnews.com/news/us-propaganda-push-in-iraq/. (accessed: 01.05.2022).

[37] *Explain Pooling layers: Max Pooling, Average Pooling, Global Average Pooling, and Global Max pooling*. URL: https://androidkt.com/explain-pooling-layers-max-pooling-average-pooling-global-average-pooling-and-global-max-pooling/. (accessed: 11.06.2022).

[38] *Fake news website*. URL: https://en.wikipedia.org/wiki/Fake_news_website. (accessed: 29.04.2022).

[39] James H Fetzer. “Disinformation: The use of false information”. In: *Minds and Machines* 14.2 (2004), pp. 231–240.

[40] Jonathan Franceschi and Lorenzo Pareschi. “Spreading of fake news, competence and learning: kinetic modelling and numerical approximation”. In: *Philosophical Transactions of the Royal Society A* 380.2224 (2022), p. 20210159.

[41] *Gated Recurrent Unit Networks*. URL: https://www.geeksforgeeks.org/gated-recurrent-unit-networks/.

[42] *Generative Adversarial Network*. URL: https://en.wikipedia.org/wiki/Generative_adversarial_network. (accessed: 07.06.2022).

[43] *Getting the Most Out of Your Google Colab (Tutorial)*. URL: https://medium.com/@oribarel/getting-the-most-out-of-your-google-colab-2b0585f82403. (accessed: 11.05.2022).

[44] Jennifer Golbeck et al. "Fake news vs satire: A dataset and analysis". In: *Proceedings of the 10th ACM Conference on Web Science*. 2018, pp. 17–21.

[45] Liang Gonog and Yimin Zhou. "A review: generative adversarial networks". In: *2019 14th IEEE conference on industrial electronics and applications (ICIEA)*. IEEE. 2019, pp. 505–510.

[46] Alex Graves, Abdel-rahman Mohamed, and Geoffrey Hinton. "Speech recognition with deep recurrent neural networks". In: *2013 IEEE international conference on acoustics, speech and signal processing*. Ieee. 2013, pp. 6645–6649.

[47] Alex Graves et al. "A novel connectionist system for unconstrained handwriting recognition". In: *IEEE transactions on pattern analysis and machine intelligence* 31.5 (2008), pp. 855–868.

[48] Sunil Gundapu and Radhika Mamidi. "Transformer based automatic COVID-19 fake news detection system". In: *arXiv preprint arXiv:2101.00180* (2021).

[49] Aditi Gupta et al. "Faking sandy: characterizing and identifying fake images on twitter during hurricane sandy". In: *Proceedings of the 22nd international conference on World Wide Web*. 2013, pp. 729–736.

[50] Sakshini Hangloo and Bhavna Arora. "Fake News Detection Tools and Methods–A Review". In: *arXiv preprint arXiv:2112.11185* (2021).

[51] Hansika Hewamalage, Christoph Bergmeir, and Kasun Bandara. "Recurrent neural networks for time series forecasting: Current status and future directions". In: *International Journal of Forecasting* 37.1 (2021), pp. 388–427.

[52] Sepp Hochreiter and Jürgen Schmidhuber. "Long short-term memory". In: *Neural computation* 9.8 (1997), pp. 1735–1780.

[53] Benjamin Horne and Sibel Adali. "This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news". In: *Proceedings of the international AAAI conference on web and social media*. Vol. 11. 1. 2017, pp. 759–766.

[54] *How The Internet Has Changed News Media Outlets*. URL: https://sites.google.com/site/newsoutletsandtheinternet/. (accessed: 24.04.2022).

[55] Kate Mathews Hunt. "Gaming the system: Fake online reviews v. consumer law". In: *Computer law & security review* 31.1 (2015), pp. 3–25.

[56] *Infodemic*. URL: https://www.who.int/health-topics/infodemic#tab=tab_1. (accessed: 24.04.2022).

[57] *Introduction to CNN*. URL: https://dev.to/deepeshaburse/introduction-to-cnn-1i03. (accessed: 11.06.2022).

[58] Md Saiful Islam et al. "COVID-19–related infodemic and its impact on public health: A global social media analysis". In: *The American journal of tropical medicine and hygiene* 103.4 (2020), p. 1621.

[59] Zhiwei Jin et al. "Novel visual and statistical image features for microblogs news verification". In: *IEEE transactions on multimedia* 19.3 (2016), pp. 598–608.

[60] *K-Nearest Neighbors Algorithm*. URL: https://www.ibm.com/topics/knn. (accessed: 08.06.2022).

[61] Bente Kalsnes. "Fake news". In: *Oxford Research Encyclopedia of Communication*. 2018.

[62] Andrea Stevens Karnyoto et al. "Transfer learning and GRU-CRF augmentation for COVID-19 fake news detection". In: *Computer Science and Information Systems* 00 (2022), pp. 53–53.

[63] By WILL KENTON. *Social Networking*. URL: https://www.investopedia.com/terms/s/social-networking.asp#:~:text=Social%5C%20networking%5C%20is%5C%20the%5C%20use,Facebook%5C%2C%5C%20Instagram%5C%2C%5C%20and%5C%20Twitter.. (accessed: 04.04.2022).

[64] *Keras*. URL: https://en.wikipedia.org/wiki/Keras. (accessed: 11.05.2022).

[65] Sagar Khillar. *Difference Between Blogging and Microblogging*. URL: http://www.differencebetween.net/technology/difference-between-blogging-and-microblogging/. (accessed: 07.06.2022).

[66] Sotiris Kotsiantis and Dimitris Kanellopoulos. "Association rules mining: A recent overview". In: *GESTS International Transactions on Computer Science and Engineering* 32.1 (2006), pp. 71–82.

[67] Srijan Kumar and Neil Shah. "False information on web and social media: A survey". In: *arXiv preprint arXiv:1804.08559* (2018). DOI: 10.48550/arXiv.1804.08559. URL: https://doi.org/10.48550/arXiv.1804.08559.

[68] Orestis Lampridis, Dimitra Karanatsiou, and Athena Vakali. "MANIFESTO: a huMAN-centric explaInable approach for FakE news spreaders deTectiOn". In: *Computing* (2022), pp. 1–23. DOI: 10.1007/s00607-021-01013-w. URL: https://doi.org/10.1007/s00607-021-01013-w.

[69] *Linear Regression*. URL: https://en.wikipedia.org/wiki/Linear_regression. (accessed: 06.06.2022).

[70] *Linear Regression for Machine Learning*. URL: https://machinelearningmastery.com/linear-regression-for-machine-learning/. (accessed: 07.06.2022).

[71] *List of fake news websites*. URL: https://en.wikipedia.org/wiki/List_of_fake_news_websites. (accessed: 26.04.2022).

[72] Zaitul Iradah Mahid, Selvakumar Manickam, and Shankar Karuppayah. "Fake news on social media: brief review on detection techniques". In: *2018 Fourth International Conference on Advances in Computing, Communication & Automation (ICACCA)*. IEEE. 2018, pp. 1–5.

[73] Alessia Mammone, Marco Turchi, and Nello Cristianini. "Support vector machines". In: *Wiley Interdisciplinary Reviews: Computational Statistics* 1.3 (2009), pp. 283–289.

[74] Maulidina Marlita. *What Is a Blog? Definition, Blog Types, and Benefits Explained*. URL: `https://www.hostinger.com/tutorials/what-is-a-blog`. (accessed: 29.04.2022).

[75] *Microblogging*. URL: `https://en.wikipedia.org/wiki/Microblogging`. (accessed: 07.06.2022).

[76] Michael Middleton. *Deep Learning vs. Machine Learning — What's the Difference?* URL: `https://flatironschool.com/blog/deep-learning-vs-machine-learning/`. (accessed: 08.02.2022).

[77] Tanushree Mitra and Eric Gilbert. "Credbank: A large-scale social media corpus with associated credibility annotations". In: *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 9. 1. 2015, pp. 258–267.

[78] Elise Moreau. *10 of the Most Popular News Blogs on the Internet*. URL: `https://www.lifewire.com/top-most-popular-blogs-3486365`. (accessed: 01.05.2022).

[79] Moin Nadeem et al. "FAKTA: An automatic end-to-end fact checking system". In: *arXiv preprint arXiv:1906.04164* (2019).

[80] Ibnu Nadzir, Sari Seftiani, and Yogi Setya Permana. *Hoax and misinformation in Indonesia: insights from a nationwide survey*. ISEAS-Yusof Ishak Institute, 2019.

[81] *Naive Bayes classifier*. URL: `https://en.wikipedia.org/wiki/Naive_Bayes_classifier`.

[82] Mohammed Amine Naji et al. "Machine Learning Algorithms For Breast Cancer Prediction And Diagnosis". In: *Procedia Computer Science* 191 (2021), pp. 487–492.

[83] *Neural Network*. URL: `https://www.investopedia.com/terms/n/neuralnetwork.asp#:~:text=A%5C%20neural%5C%20network%5C%20is%5C%20a,organic%5C%20or%5C%20artificial%5C%20in%5C%20nature`. (accessed: 27.04.2022).

[84] *Neural Networks*. URL: `https://www.ibm.com/cloud/learn/neural-networks`.

[85] Giang Nguyen et al. "Machine learning and deep learning frameworks and libraries for large-scale data mining: a survey". In: *Artificial Intelligence Review* 52.1 (2019), pp. 77–124. DOI: `10.1007/s10462-018-09679-z`. URL: `https://doi.org/10.1007/s10462-018-09679-z`.

[86] Keiron O'Shea and Ryan Nash. "An introduction to convolutional neural networks". In: *arXiv preprint arXiv:1511.08458* (2015).

[87] Prathmesh Pathwar and Simran Gill. "Tackling COVID-19 infodemic using deep learning". In: *arXiv preprint arXiv:2107.02012* (2021).

[88] Parth Patwa et al. "Fighting an infodemic: Covid-19 fake news dataset". In: *International Workshop on Combating On line Ho st ile Posts in Regional Languages dur ing Emerge ncy Si tuation*. Springer. 2021, pp. 21–29.

[89] Nathaniel Persily and Joshua A Tucker. *Social Media and Democracy: The State of the Field, Prospects for Reform*. Cambridge University Press, 2020.

[90] Leif E Peterson. "K-nearest neighbor". In: *Scholarpedia* 4.2 (2009), p. 1883.

[91] Michael Phi. *Illustrated Guide to LSTM's and GRU's: A step by step explanation*. URL: `https://towardsdatascience.com/illustrated-guide-to-lstms-and-gru-s-a-step-by-step-explanation-44e9eb85bf21`. (accessed: 08.06.2022).

[92] Wenjing Pian, Jianxing Chi, and Feicheng Ma. "The causes, impacts and countermeasures of COVID-19 "Infodemic": A systematic review using narrative synthesis". In: *Information processing & management* 58.6 (2021), p. 102713.

[93] *Political bias on social media emerges from users, not platform, IU study says*. URL: `https://research.impact.iu.edu/key-areas/social-sciences/stories/social-media-platform-bias.html`. (accessed: 28.04.2022).

[94] Martin Potthast et al. "A stylometric inquiry into hyperpartisan and fake news". In: *arXiv preprint arXiv:1702.05638* (2017).

[95] Cristina M Pulido et al. "A new application of social impact in social media for overcoming fake news in health". In: *International journal of environmental research and public health* 17.7 (2020), p. 2430.

[96] *Python(programming$_l$anguage)*. URL: `https://en.wikipedia.org/wiki/Python_(programming_language)`. (accessed: 11.05.2022).

[97] *QA software*. URL: `https://en.wikipedia.org/wiki/Q%5C%26A_software`. (accessed: 05.06.2022).

[98] Thorsten Quandt et al. "Fake news". In: *The international encyclopedia of journalism studies* (2019), pp. 1–6.

[99] Sunil Ray. *Understanding Support Vector Machine(SVM) algorithm from examples (along with code)*. URL: `https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/`. (accessed: 08.06.2022).

[100] Shaina Raza and Chen Ding. "Fake news detection based on news content and social contexts: a transformer-based approach". In: *International Journal of Data Science and Analytics* (2022), pp. 1–28. DOI: `10.1007/s41060-021-00302-z`. URL: `https://doi.org/10.1007/s41060-021-00302-z`.

[101] *Review site*. URL: `https://en.wikipedia.org/wiki/Review_site`. (accessed: 05.06.2022).

[102] Steven J Rigatti. “Random forest”. In: *Journal of Insurance Medicine* 47.1 (2017), pp. 31–39.

[103] Lior Rokach and Oded Maimon. “Decision trees”. In: *Data mining and knowledge discovery handbook*. Springer, 2005, pp. 165–192.

[104] Sharat Sachin et al. “Sentiment analysis using gated recurrent neural networks”. In: *SN Computer Science* 1.2 (2020), pp. 1–13.

[105] SM Sadiq-Ur-Rahman Shifath, Mohammad Faiyaz Khan, and Md Saiful Islam. “A transformer based approach for fighting COVID-19 fake news”. In: *arXiv e-prints* (2021), arXiv–2101.

[106] Paul Sagan and Frank (Tom) Thomson Leighton. *The Internet & the future of news*. URL: https://www.amacad.org/publication/internet-future-news. (accessed: 24.04.2022).

[107] Sagar Shamra. *Activation Functions in Neural Networks*. URL: https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6. (accessed: 27.02.2022).

[108] Scott Shane. *The Fake Americans Russia Created to Influence the Election*. URL: https://www.nytimes.com/2017/09/07/us/politics/russia-facebook-twitter-election.html. (accessed: 30.04.2022).

[109] Kai Shu, Suhang Wang, and Huan Liu. “Exploiting tri-relationship for fake news detection”. In: *arXiv preprint arXiv:1712.07709* 8 (2017).

[110] Kai Shu et al. “Fake news detection on social media: A data mining perspective”. In: *ACM SIGKDD explorations newsletter* 19.1 (2017), pp. 22–36.

[111] Elena Shushkevich and John Cardiff. “TUDublin team at Constraint@ AAAI2021–COVID19 Fake News Detection”. In: *arXiv preprint arXiv:2101.05701* (2021).

[112] Kristina P Sinaga and Miin-Shen Yang. “Unsupervised K-means clustering algorithm”. In: *IEEE access* 8 (2020), pp. 80716–80727.

[113] *Sock puppet account*. URL: https://en.wikipedia.org/wiki/Sock_puppet_account. (accessed: 01.05.2022).

[114] Carlos Oscar Sánchez Sorzano, Javier Vargas, and A Pascual Montano. “A survey of dimensionality reduction techniques”. In: *arXiv preprint arXiv:1403.2877* (2014).

[115] Xiaogang Su, Xin Yan, and Chih-Ling Tsai. “Linear regression”. In: *Wiley Interdisciplinary Reviews: Computational Statistics* 4.3 (2012), pp. 275–294.

[116] Edson C Tandoc Jr, Zheng Wei Lim, and Richard Ling. “Defining “fake news” A typology of scholarly definitions”. In: *Digital journalism* 6.2 (2018), pp. 137–153.

[117] Python Core Team. “Python: A dynamic, open source programming language”. In: *Python Software Foundation* 78 (2015).

[118] Gerald Tesauro et al. “Temporal difference learning and TD-Gammon”. In: *Communications of the ACM* 38.3 (1995), pp. 58–68.

[119] *The Death of President Bouteflika just a hoax- official*. URL: https://www.moroccoworldnews.com/2012/09/54914/the-death-of-president-bouteflika-just-a-hoax-official. (accessed: 28.04.2022).

[120] *Top 10 Deep Learning Algorithms You Should Know in 2022*. URL: https://www.simplilearn.com/tutorials/deep-learning-tutorial/deep-learning-algorithm#defining_neural_networks. (accessed: 07.06.2022).

[121] *Understanding Generative Adversarial Networks (GANs)*. URL: https://towardsdatascience.com/understanding-generative-adversarial-networks-gans-cd6e4651a29. (accessed: 07.06.2022).

[122] *Understanding GRU Networks*. URL: https://towardsdatascience.com/understanding-gru-networks-2ef37df6c9be.

[123] *Understanding K-means Clustering in Machine Learning*. URL: https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1. (accessed: 08.06.2022).

[124] *Understanding Random Forest*. URL: https://towardsdatascience.com/understanding-random-forest-58381e0602d. (accessed: 08.06.2022).

[125] Ike Vayansky and Sathish AP Kumar. "A review of topic modeling methods". In: *Information Systems* 94 (2020), p. 101582.

[126] Andreas Vlachos and Sebastian Riedel. "Fact checking: Task definition and dataset construction". In: *Proceedings of the ACL 2014 workshop on language technologies and computational social science*. 2014, pp. 18–22.

[127] Chih-Chien Wang. "Fake news and related concepts: Definitions and recent research development". In: *Contemporary Management Research* 16.3 (2020), pp. 145–174.

[128] Apurva Wani et al. "Evaluating deep learning approaches for covid19 fake news detection". In: *International Workshop on Combating On line Ho st ile Posts in Regional Languages dur ing Emerge ncy Si tuation*. Springer. 2021, pp. 153–163.

[129] Christopher JCH Watkins and Peter Dayan. "Q-learning". In: *Machine learning* 8.3 (1992), pp. 279–292.

[130] *Web Forum*. URL: https://techterms.com/definition/web_forum. (accessed: 26.04.2022).

[131] Geoffrey I Webb, Eamonn Keogh, and Risto Miikkulainen. "Naıve Bayes." In: *Encyclopedia of machine learning* 15 (2010), pp. 713–714.

[132] *What are social news websites?* URL: https://blogs.bmj.com/bmj-journals-development-blog/tag/social-news/. (accessed: 26.04.2022).

[133] Thomas Wood. *Backpropagation*. URL: https://deepai.org/machine-learning-glossary-and-terms/backpropagation. (accessed: 23.05.2022).

[134] Yonghui Wu et al. "Google's neural machine translation system: Bridging the gap between human and machine translation". In: *arXiv preprint arXiv:1609.08144* (2016).

[135] Yuanyuan Wu et al. "Fake online reviews: Literature review, synthesis, and directions for future research". In: *Decision Support Systems* 132 (2020), p. 113280. ISSN: 0167-9236. DOI: `https://doi.org/10.1016/j.dss.2020.113280`. URL: `https://www.sciencedirect.com/science/article/pii/S016792362030035X`.

[136] Jyoti Yadav and Monika Sharma. "A Review of K-mean Algorithm". In: *Int. J. Eng. Trends Technol* 4.7 (2013), pp. 2972–2976.

[137] Savvas Zannettou et al. "The web of false information: Rumors, fake news, hoaxes, clickbait, and various other shenanigans". In: *Journal of Data and Information Quality (JDIQ)* 11.3 (2019), pp. 1–37. DOI: `10.1145/3309699`. URL: `https://doi.org/10.1145/3309699`.

[138] Yagang Zhang. *New Advances in Machine Learning*. Rijeka: IntechOpen, 2010. DOI: `10.5772/225`. URL: `https://doi.org/10.5772/225`.

[139] Xinyi Zhou and Reza Zafarani. "A survey of fake news: Fundamental theories, detection methods, and opportunities". In: *ACM Computing Surveys (CSUR)* 53.5 (2020), pp. 1–40. DOI: `10.1145/3395046`. URL: `https://doi.org/10.1145/3395046`.

[140] Zhi-Hua Zhou. *Ensemble methods: foundations and algorithms*. CRC press, 2012.

[141] Arkaitz Zubiaga et al. "Detection and Resolution of Rumours in Social Media: A Survey". In: *ACM Comput. Surv.* 51.2 (Feb. 2018). ISSN: 0360-0300. DOI: `10.1145/3161603`. URL: `https://doi.org/10.1145/3161603`.