

**FACULTY OF COMPUTER SCIENCE & INFORMATION TECHNOLOGY**



RESEARCH PROPOSAL

**Text analytics on deceptive and fake information on COVID-19 in Malaysia**

CHEN BAOGANG

17186722/2

Supervisor: Dr. Siti Soraya binti Abdul Rahman

Contents

[1. Abstract 3](#_Toc135254853)

[2. Introduction 3](#_Toc135254854)

[3. Problem statement 4](#_Toc135254855)

[4. Research questions: 5](#_Toc135254856)

[5. Research objectives: 6](#_Toc135254857)

[6. Research Scope and Contribution : 6](#_Toc135254858)

[7. Related work 7](#_Toc135254859)

[8. Methodology: 10](#_Toc135254860)

[9. Result 14](#_Toc135254861)

[10. DISCUSSIONS 14](#_Toc135254862)

[11. Conclusion 14](#_Toc135254863)

[12. IMPLICATIONS OF RESEARCH 14](#_Toc135254864)

[13. LIMITATIONS AND FUTURE RESEARCH 14](#_Toc135254865)

[14. ACKNOWLEDGEMENT 14](#_Toc135254866)

[15. Reference: 14](#_Toc135254867)

# Abstract

The COVID-19 pandemic has witnessed the proliferation of fake news, which poses a significant threat by potentially leading individuals to disregard necessary precautions for their own safety and that of others or to believe in unproven treatments and cures. This study aims to address the issue of fake news about COVID-19 and prevent its detrimental effects. Specifically, it investigates deceptive and fake information related to COVID-19 and proposes the utilization of machine learning models for detection purposes. The primary objectives of this research are to evaluate various machine learning models based on their performance metrics and select the most effective prediction model. To achieve this, a diverse dataset from multiple sources will be collected, and six machine learning algorithms—Decision Trees, Logistic Regression, Gradient Boost, k-Nearest Neighbors (KNN), Naive Bayes (NB), and Support Vector Machines (SVM)—will be employed for analysis. By implementing the proposed framework, it is expected that the detection of fake and non-fake tweets containing COVID-19 information will yield highly accurate results. Ultimately, this research aims to contribute to the mitigation of fake news and its impact during the COVID-19 pandemic.

# Introduction

In December 2019, the World Health Organization received notification about a group of pneumonia cases with an unknown etiology that were detected in Wuhan city and rapidly spread throughout China, and the patients who were affected by the unknown pneumonia cases were reported to have specific symptoms, including a dry cough, difficulty breathing, fever, and signs of inflammation in both lungs as seen on imaging scans. The virus responsible for this outbreak was later identified as the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) and the illness caused by the virus was named COVID-19 by the World Health Organization in 2020 (World Health Organization,2020).

The rapid spread of this virus was facilitated by modern transportation and social media. Because advancements in transportation and the widespread use of social media have made the world feel smaller, as it is now easier to move people and information from one place to another. This has led to the rapid spread of COVID-19, making it a pandemic. Social media platforms also play a significant role in the spread of information, both accurate and misleading, about the virus and its impact. The presence of a new virus whose characteristics are not fully understood, combined with fear and panic among the public, has led to the circulation of a lot of false information about the virus. (Elhadad et al., 2020)

Deceptive and fake information usually refer to any false or inaccurate information that is spread deliberately or unintentionally. It can take many forms, such as rumors, fake news, or misleading content on the internet. Deceptive and fake information can be spread through various channels, but primarily social media, due to the Internet's and related technologies' rapid development, social media and web usage are becoming increasingly popular over time. More than 3.6 billion people are using social media as of 2020, and by 2025, it's predicted the number would reach 4.41 billion (J.Cement, 2020).

However, the existence of deceptive and fake information about the COVID-19 also grows on the internet, unreliable and false information is spreading around the world to such an extent, that some commentators are now referring to the new avalanche of deceptive and fake information that’s accompanied the COVID-19 pandemic as a ‘disinfodemic’ (United Nations,2020).

The spread of deceptive and fake information can cause confusion, mistrust, and even be used to manipulate public opinion and decision-making, with serious consequences for individuals, communities, and society (Domenico et al., 2021). This is particularly concerning during the COVID-19 pandemic, as people rely heavily on news and social media for information to make decisions about their safety. Moreover, deceptive and fake information about COVID-19 can lead to people not taking necessary precautions to protect themselves and others or believing in unproven treatments and cures. Furthermore, this can also present challenges for healthcare workers and disrupt society, leading to negative effects on culture, economics, and reducing people's confidence in their governments.

Therefore, it is important to study existing technologies for detecting misleading information in order to combat deceptive and fake information. There are some organizations and fact-checkers that have been created to combat deceptive and fake information and promote accurate information, they use various techniques such as machine learning, fact-checking by journalists and subject-matter experts, and legal action against deceptive and fake information spreaders. In this research project, the goal is to examine existing misleading-information detection models, compare the implementation of different machine learning and deep learning models, and evaluate the performance of various technologies. The ultimate objective is to create an efficient detecting deceptive and fake information system that obtains news or articles about covid-19 from different resources.

# Problem statement

The COVID-19 pandemic has raised concerns about the spread of deceptive false information, particularly through social media and the internet. This can lead to dangerous beliefs about masks and vaccines, as well as disregard for public health measures (Naeem et al., 2021), causing widespread mistrust of official health messages and the healthcare system as a whole (do Nascimento et al., 2022). The issue of identifying deceptive and fake information is difficult due to its various forms, including disinformation, misinformation, hoaxes, propaganda, satire, rumors, and click-bait (Kudarvalli & Fiaidhi,2020). Allowing deceptive and fake information to spread can have dangerous consequences, as seen in Iran where a false claim about alcohol curing COVID-19 led to hospitalizations and fatalities (Karimi and Gambrell, 2020). The problem of deceptive and fake information requires a solution, despite efforts to create systems to detect it, most still rely on manually labeled data (Elhadad et al., 2020) and only focus on one dataset and social media platform -Twitter (Alenezi & Alqenaei,2021). Further research is necessary to address this issue by investigating multiple datasets and developing a reliable and accurate model to detect deceptive and fake information on COVID-19.

# Research questions:

* What are the impacts of deceptive and fake information on COVID-19 on individuals and society as a whole?
* What are the different algorithms and techniques used in machine learning to detect deceptive and fake information about COVID-19?
* What are the performance metrics that can be used to evaluate the effectiveness of the machine learning models in detecting deceptive and fake information about COVID-19?
* What are the different types and forms of deceptive and fake information related to COVID-19, and how are they spread through social media and other channels?
* How do individuals perceive and react to deceptive and fake information related to COVID-19, and what are the potential impacts on their beliefs, attitudes, and behaviors towards public health measures?
* How does the spread of deceptive and fake information related to COVID-19 affect the healthcare system, government policies, and public trust in official sources of information?
* What methods and techniques are currently available to detect and combat deceptive and fake information related to COVID-19, and what are their limitations and challenges?
* How can machine learning and natural language processing techniques be used to develop a reliable and accurate model to detect deceptive and fake information related to COVID-19, using multiple datasets and social media platforms?
* What datasets related to COVID-19 are currently available for detecting and combating deceptive and fake information, and what are their characteristics, strengths, and limitations in terms of volume, diversity, and quality of information?
* To investigate and analyze existing datasets related to COVID-19 from diverse sources.

# Research objectives:

* To investigate deceptive and fake information on COVID-19
* To build machine learning model to detect deceptive and fake information on COVID-19
* To evaluate several machine learning models on performance metrics and choose the best prediction model.

# Research Scope and Contribution:

Research Scope:

This study focuses on investigating and analyzing existing datasets related to COVID-19 with the aim of detecting and combating deceptive and fake information. The research will encompass diverse datasets collected from various sources, such as social media platforms, news articles, official health organizations, research institutions, and public databases. The analysis will primarily involve the application of the "OSMN" framework to obtain, scrub, model, and interpret the data.

Research Contribution:

The findings of this research will contribute to the understanding of deceptive and fake information surrounding the COVID-19 pandemic. By analyzing diverse datasets, the study aims to provide insights into the prevalence, characteristics, and patterns of deceptive and fake information related to COVID-19. This knowledge can be utilized to develop effective detection models and strategies to combat misinformation, enhancing public health communication and decision-making during the ongoing pandemic.

Furthermore, this research aims to contribute to the existing body of literature on the topic by evaluating the performance of the "OSMN" framework in the context of analyzing COVID-19 datasets for detecting deceptive and fake information. The application of the framework will provide practical insights into the data analysis process and its effectiveness in addressing the research problem.

Overall, the research aims to provide valuable insights into the detection and mitigation of deceptive and fake information related to COVID-19, ultimately contributing to the broader goals of promoting accurate information, protecting public health, and fostering trust in reliable sources during challenging times.

# Related work

# 1 introduction

In recent years, the proliferation of fake news has emerged as a significant concern due to its rapid spread and consequential real-world impact. Conventional supervised machine learning techniques have played a pivotal role in combating this issue. However, with the advent of deep learning, there has been a paradigm shift in the field of text classification, offering promising advancements in the detection of fake news.

In light of these advancements, this literature review aims to explore the current landscape of covid 19 fake news detection, comparing and contrasting the effectiveness of conventional supervised machine learning techniques and deep learning approaches.

# Literature review

Fake news has become a major problem in recent years, as it can spread quickly and have significant real-world consequences, but conventional supervised machine learning techniques have been quite helpful in addressing this issue,

In 2020, (Patwa et al., n.d.) proposed a manually annotated dataset of 10,700 social media posts and articles of real and fake news on COVID-19, including but not limited to tweets, Facebook and Instagram posts, public statements, and other popular media content, additionally, the fake news relate to COVID-19 is verified on well-known fact-checking websites like PolitiFact, Snopes, Boomlive, NewsChecker and from other tools like IFCN chatbot and Google fact-check-explorer. While real tweets are scraped from official government accounts, medical institutions, news channels, etc. via the twitter API, and read by humans if it contains useful information (e.g., numbers, dates, vaccine progress, government policies, hotspots) about COVID-19 are marked as real news, therefore, the problem is then treated as a binary classification problem where these data are fed into conventional Machine Learning classifiers like Decision Tree, Logistic Regression, Gradient Boost, and Support Vector Machine(SVM), where SVM performs the best F1-score with 93.32%.

Reis et al., (2019) use a dataset consisting of 2282 BuzzFeedNews articles about the 2016 U.S. election that have been labeled by journalists, and categorize features for fake news detection into three levels, such as textual feature, news source features and environment feature, and then, using these hand-craft features as input to conventional machine learning classifiers , including k-Nearest Neighbors (KNN), Naive Bayes (NB), Random Forests(RF), Support Vector Machine (SVM), and XGBoost (XGB), where the best performance was obtained statistically by the RF and XGB, with AUC values of 0.85 (0.007) and 0.86 (0.006), respectively.

Liu et al. (2023) propose the development of an online software detector that incorporates various machine learning techniques, including traditional machine learning, ensemble machine learning, and deep learning mechanisms. To prepare the input dataset, the authors apply preprocessing techniques such as removing redundant information, tokenization, stopword removal, and lemmatization. They then employ feature extraction to convert the text-based documents into meaningful numerical representations. The authors compare the performance of different models and find that the BERT model achieves the highest accuracy of 94.4%, F-1 score of 94.4%, precision of 95%, and recall of 93.6%. Consequently, the BERT model is selected as the optimal choice for the online detector.

Z. Tian et al., (2021) suggests that machine learning can effectively detect fake news. The authors propose a fake news detection system using a k-nearest neighbors (KNN) machine learning model. By incorporating Genetic and Evolutionary Feature Selection (GEFeS) in the system, they achieve a high accuracy of 91.3%. Furthermore, they explore the application of quantum machine learning techniques in fake news detection by using the identified GEFeS features and an optimal k value to train and test a quantum KNN (QKNN) model. The QKNN model achieves an accuracy of 84.4%.

R. Malhotra et al., (2022) utilize a dataset consisting of 10,700 tweets and online posts containing fake and real news about COVID-19 obtained from ConstraintAI'21. They compare the performance of various machine learning (ML) algorithms including Logistic Regression, K-Nearest Neighbor, Linear Support Vector Machine, Random Forest Classifier, Decision Tree, Naive Bayes, and Stochastic Gradient Descent, as well as deep learning (DL) algorithms such as Recurrent Neural Network, Long Short-Term Memory, and Gated Recurrent Unit. Two feature extraction techniques, count vectorization and TF-IDF, are used. The results indicate that the Linear Support Vector Machine model with TF-IDF achieves the highest testing accuracy of 94.11%, closely followed by the Stochastic Gradient Descent classifier with 93.92%.

Bang et al., (2021) attempted to create a strong model for identifying fake news that would work well across different test sets. They tested their model on two different datasets, FakeNews-19, and Tweets-19, and found that while one approach, which involved fine-tuning transformer-based language models with robust loss functions, did not greatly improve performance on the FakeNews-19 dataset, it did show better generalization on the Tweets-19 dataset. They also implemented an influence-based data cleansing technique which made the model more robust and adaptable.

Abdelminaam et al., (2021) presents a framework called CoAID-DEEP for detecting deceptive and fake information about COVID-19 on Twitter. The authors used six machine learning models (DT, LR, KNN, RF, SVM, NB) and Two deep learning models (Modified LSTM, Modified GRU) for the detection of deceptive and false information, the authors report that the framework achieved high accuracy scores on the COVID-19 tweets dataset, outperforming other state-of-the-art models in terms of both accuracy and computational efficiency.

In end-user environment, (Sahoo & Gupta, 2021) proposed a fake news detection framework to detect fake news in Facebook platform on user’s home page using combination of machine learning (KNN,SVM Logistic Regression, Decision tree, Naïve Bayes ) and deep learning (Long Short-term Memory (LSTM)), in this study, authors analyzes both user profile (Profile ID, name, Date of join, All friends etc..) and news content features (Source, Headline, Body text, Statistical features etc..), the accuracy of the detection system scores 99.4%, the excellent performance shows that multiple features related to the user account and shared content in the form of fake news can significantly improve the accuracy of fake news detection.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Citations | Article | Data & features | Models | Analysis & Results |
| Patwa et al., n.d. (2020) | Fighting an Infodemic: COVID-19 Fake News Dataset | 10,700 social media posts, numbers, dates, vaccine progress, government policies, hotspots | Decision Tree, Logistic Regression, Gradient Boost, and Support Vector Machine (SVM) | SVM performs the best F1-score with 93.32%. |
| Reis et al., (2019) | Supervised Learning for Fake News Detection | 2282 BuzzFeedNews articles, three levels: textual feature, news source features and environment feature | k-Nearest Neighbors (KNN), Naive Bayes (NB), Random Forests(RF), Support Vector Machine (SVM), and XGBoost (XGB) | RF and XGB, with AUC values of 0.85 (0.007) and 0.86 (0.006), respectively |
| Das et al., 2021 | A Heuristic-driven Ensemble Framework for COVID-19 Fake News Detection | 10,700 social media news, three levels: Article body, Username, domain | pre-trained models, like XLNet , RoBERTa , XLM-RoBERTa , DeBERTa , ERNIE 2.0 and ELECTRA models | best performing ensemble model consisting of RoBERTa, XLM-RoBERTa, XLNet , DeBERTa and novel heuristic algorithm based on username handles and link domains in tweets fetching an F1-score of 0.9883 |
| Sahoo & Gupta, 2021 | Multiple features based approach for automatic fake news detection on social networks using deep learning | More than 15000 news, user profile (Profile ID, name, Date of join, All friends etc..) and news content features (Source, Headline, Body text, Statistical features etc..) | Machine learning classifiers: KNN, SVM LR, DT, NB  Deep learning classifiers (LSTM) | Accuracy KNN=99.3 SVM=99.3 Logistic regression=99.0 Decision tree=99.1 LSTM=99.4 |
| Abdelminaam et al., 2021 | CoAID-DEEP: An Optimized Intelligent Framework for Automated Detecting COVID-19 Misleading Information on Twitter | CoAID dataset (Id, text, location,keyword, target | feature extraction by TF-IDF, DT, RF, KNN, SVM, LR, NB, Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU) | The Modified LSTM and The Modified GRU for the gossip cop dataset are outperforming SVM, DT,RF, NB, LR, and KNN for cross-validation and testing results, LSTM accuracy= 98.6% |
| J. Liu et al., (2023) | COVID-19 Fake News Detector | 10, 700 English posts | DT, RF, KNN, SVM, LR, GBC, NB, BERT, RoBERTa | BERT performs the best accuracy of 94.4% |
| Z. Tian et al., (2021) | Fake News Detection using Machine Learning with Feature Selection | BuzzFace consists of 2282 news articles | KNN, Genetic and Evolutionary Feature Selection (GEFeS) | With GETeS , KNN achieve a high accuracy of 91.3%. |
| R. Malhotra et al.,(2022) | COVID-19 Fake News Detection System | 10,700 tweets and online posts | feature extraction by TF-IDF, count vectorization  ML models: DT, RF, KNN, SVM, LR, NB, LSVM, SGD  DL models: LSTM, GRU, RNN | LSVM with TF-IDF achieves the highest testing accuracy of 94.11% |
| N. L. S. R. Krishna et al., (2022) | Fake News Detection system using Decision Tree algorithm and compare textual property with Support Vector Machine algorithm |  |  |  |
|  |  |  |  |  |

# 3 Research gap

# A research gap exists in the investigation of feature selection techniques in the context of fake news detection. Feature selection plays a crucial role in improving the performance and interpretability of machine learning models. However, most researchers tend to focus on a single type of feature extraction, such as TF-IDF, and often overlook the exploration of other feature extraction techniques or the comparison of their effectiveness.

# While some studies, like the work by Z. Tian et al. (2021), have applied feature selection techniques like Genetic and Evolutionary Feature Selection (GEFeS), these methods have not been widely adopted in the field of fake news detection. The lack of broader adoption limits the understanding of the potential benefits that different feature selection approaches can bring to the performance of fake news detection models.

# Future research could address this gap by investigating the effectiveness of different feature selection approaches and exploring their impact on model performance. By comparing the accuracy and interpretability of models using various feature selection techniques, researchers can gain insights into the most effective methods for identifying informative and relevant features in the context of fake news detection. This would contribute to the development of more robust and accurate detection systems that can effectively combat the spread of deceptive information.

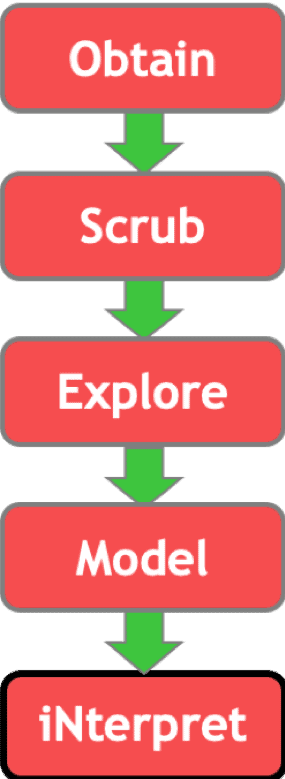
# Summary

In summary, this literature review examines the effectiveness of machine learning and deep learning techniques in detecting deceptive and fake information. Deep learning methods consistently outperform machine learning methods in this regard. The review highlights the excellent results achieved by six conventional machine learning models, namely Decision Trees, Logistic Regression, Gradient Boost, k-Nearest Neighbors (KNN), Naive Bayes (NB), and Support Vector Machine (SVM). Additionally, feature extraction techniques such as TF-IDF and count vectorization are widely utilized in these studies. The review compares the performance of various machine learning models and deep learning models, with the aim of providing a theoretical foundation for future research in the field of predicting COVID-19 fake news. The findings offer valuable insights for selecting accurate prediction models in future studies.

# Methodology:

* 1. **A proposed data analysis framework**

Data analysis can pose a challenge, particularly when it comes to knowing where to begin. Even if you are familiar with some basic data analysis techniques, it can be difficult to determine the initial steps to take and what should come next. A data analysis framework can be useful in addressing this issue by offering a straightforward, step-by-step process to follow. It gives you a framework to work within and helps you to break down the complex task of data analysis into more manageable steps. Among the various standard frameworks for data science project life cycles, OSEMS is the most widely accepted. The OSEMN framework is an acronym that defines each step of the data analysis process, it includes Obtain, Scrub, Explore, Model the data, and interpret the data, in short, OSEMN’s five steps are described below:



**Obtain data:** we collect Malaysia COVID-19 related news, information, and articles from web search engines like google, Bing, yahoo, etc. Below are the datasets of various deceptive and fake information related to covid-19 spread through social media and online platforms.

* COVID-19 Fake News
* CoAID: COVID-19 Healthcare Deceptive and Fake information Dataset
* ReCOVery
* COVID-FAKES

**Scrub data**: Data scrubbing is the process of modifying or removing incomplete, inaccurate, poorly formatted, or duplicate data in a database. It is also well known as data Pre-processing procedures in data analysis, in this research include fundamental cleaning operations including low casing, eliminating unnecessary characters, Tokenizing, stop-word removal and feature extraction.

* Lower Casing: Converting all text to lowercase. This helps to standardize the text and reduce the number of unique words that need to be analyzed. For example, "DETECTION" and "detection" would be treated as the same word.
* Removing unimportant data: Clearing data from invalid values, commas, apostrophes, quotes, question marks or any non-English letters, symbols, or non-alphanumeric characters, and only retaining English letters, numbers, or a combination of both.
* Tokenizing: Tokenization is the process of breaking up a sentence or paragraph into individual words or units, known as tokens. For example, the sentence "detection system" would be broken into the tokens "detection", and "system" after tokenization.
* Stop-word: A stop word is a commonly used word in a language that does not contribute significant meaning to a sentence, such as articles, prepositions, conjunctions, and some pronouns. For example, the words “the”, “a”, “and” ,“is” will be identified as stop-words.
* Steamming: by using the root word of each word, which helps avoid repetitive patterns. For example, the words “running”, ‘‘ran,’’ and ‘‘runs’’ will be reduced to the word “run”.

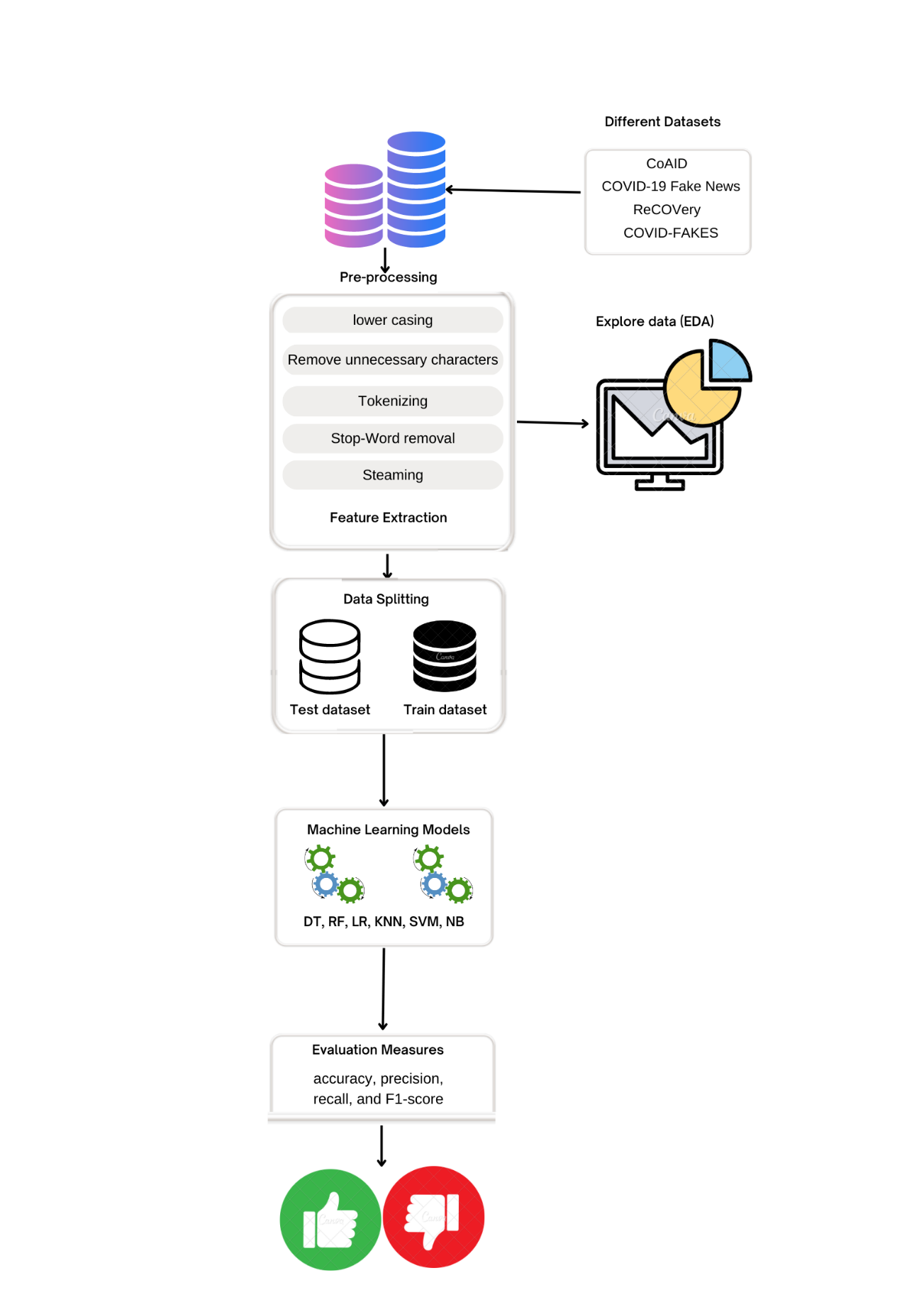
**Explore data (EDA):** Explore the data called exploratory data analysis.

**Model data:** Regarding models, the summary of the literature review highlights the use of conventional ML models such as Decision Tree, Logistic Regression, Gradient Boost, Support Vector Machine (SVM), k-Nearest Neighbors (KNN), Naive Bayes (NB), Random Forests (RF), and XGBoost (XGB) will perform the best results. In this step, 6 machine Learning models (decision trees, logistic regression, k nearest neighbors, random forests, support vector machines, and naïve Bayes) are chosen based on literature review.

**Interpret the data:** The interpretation step, also known as the evaluation or examination phase, is an essential part of the OSEMN model. The main goal of this step is to draw insights and conclusions from the data that was obtained, cleaned, and analyzed in the previous steps. The interpretation step is used to answer the research questions that motivated the data modeling process, and to provide insights that can be used to inform decision-making or further research. In this step, it is important to use a variety of evaluation measures to assess the performance of different models for building a deceptive and fake information detection system. Some commonly used evaluation measures for this type of system include accuracy, precision, recall, and F1-score. This step includes:

* Drawing conclusions from the datasets.
* Evaluation of different models.
* Select suitable models to build a deceptive and fake information detection system.
  1. **The** **deceptive and fake information detection system design**

The diagram presented illustrates the overall design of the system for detecting misleading information about covid-19. The method for identifying misleading information is broken down into various stages that align with the OMEMN framework.



From engineering perspective, most of the current fact checker are not domain specific so fail to performance well for COVID domain, while our system is among the first online COVID fake news detector websites to fill this gap to fight the wide spreading of COVID related fake news

# Result

# DISCUSSIONS

# Conclusion

# IMPLICATIONS OF RESEARCH

# LIMITATIONS AND FUTURE RESEARCH

A limitation of this fake news detection research is that it mainly focuses on text-based features and content analysis, leaving gaps in exploring the potential of user profiles and behaviors as informative indicators. This research mainly revolves around textual information and does not fully investigate the impact of other relevant features on fake news prediction systems. For example, factors such as the source of the article, the field of source, and the credibility of the author may all affect the authenticity of the news. Incorporating these additional features into detection models can improve the accuracy and robustness of fake news detection systems.

In addition to the limited consideration of user-related features, the research mainly focuses on machine learning models, and does not fully study deep learning models. Future work can include research on deep learning architectures: deep learning models, such as LSTM, GRU, and BERT, In addition to this, one can also focus on user-related features, such as user engagement, social network structure, temporal patterns, and further, besides using text as input for fake news detection, other features can also be considered, such as news source, domain and news writers, as an additional feature to improve prediction accuracy.

# ACKNOWLEDGEMENT

# Reference:

Abdelminaam, D. S., Ismail, F. H., Taha, M., Taha, A., Houssein, E. H., & Nabil, A. (2021). CoAID-deep: An optimized intelligent framework for automated detecting COVID-19 misleading information on Twitter. *IEEE Access*, *9*, 27840–27867. https://doi.org/10.1109/access.2021.3058066

Alenezi, M. N., & Alqenaei, Z. M. (2021). Machine learning in detecting COVID-19 misinformation on Twitter. *Future Internet*, *13*(10), 244. <https://doi.org/10.3390/fi13100244>

Bang, Y., Ishii, E., Cahyawijaya, S., Ji, Z., & Fung, P. (2021). *Model Generalization on*

*COVID-19 Fake News Detection*. <http://arxiv.org/abs/2101.03841>

Das, S. D., Basak, A., & Dutta, S. (2021a). *A Heuristic-driven Ensemble Framework for COVID-19 Fake News Detection*. <http://arxiv.org/abs/2101.03545>

Das, S. D., Basak, A., & Dutta, S. (2021b). *A Heuristic-driven Uncertainty based Ensemble Framework for Fake News Detection in Tweets and News Articles*. <http://arxiv.org/abs/2104.01791>

do Nascimento, I. J. B., Pizarro, A. B., Almeida, J. M., Azzopardi-Muscat, N., Gonçalves, M. A., Björklund, M., & Novillo-Ortiz, D. (2022). Infodemics and health misinformation: a systematic review of reviews. In *Bulletin of the World Health Organization* (Vol. 100, Issue 9, pp. 544–561). World Health Organization. <https://doi.org/10.2471/BLT.21.287654>

Domenico, G. di, Sit, J., Ishizaka, A., & Nunan, D. (2021). Fake news, social media and marketing: A systematic review. *Journal of Business Research*, *124*, 329–341. <https://doi.org/10.1016/J.JBUSRES.2020.11.037>

Elhadad, M. K., Li, K. F., & Gebali, F. (2020). Detecting misleading information on COVID-19. IEEE Access, 8, 165201–165215. <https://doi.org/10.1109/ACCESS.2020.3022867>

J.Cement. 2020. Number of social media users 2025. Statista. Accessed: 2020-10-30.

Kudarvalli, H., & Fiaidhi, J. (2020). Detecting fake news using machine learning algorithms.

<https://doi.org/10.36227/techrxiv.12089133>

Naeem, S. bin, Bhatti, R., & Khan, A. (2021). An exploration of how fake news is taking over social media and putting public health at risk. In *Health Information and Libraries Journal* (Vol. 38, Issue 2, pp. 143–149). Blackwell Publishing Ltd. https://doi.org/10.1111/hir.12320

Nasser Karimi and Jon Gambrell. 2020. Hundreds die of poisoning in iran as fake news

suggests methanol cure for virus. Times of Israe

Palosky, C. (2021). *Covid-19 misinformation is ubiquitous: 78% of the public believes or is*

*unsure about at least one false statement, and nearly a third believe at least four of eight false statements tested*. KFF. Retrieved November 1, 2022, from https:/[/www.kff.org/coronavirus](http://www.kff.org/coronavirus-covid-19/press-release/covid-19-misinformation-is-)-[covid-19/press-release/covid-19-misinformation-is-](http://www.kff.org/coronavirus-covid-19/press-release/covid-19-misinformation-is-) ubiquitous-78-of-the-public-believes-or-is-unsure-about-at-least-one-false-statement- and-nearly-at-third-believe-at-least-four-of-eight-false-statements-tested/

Patwa, P., Sharma, S., Pykl, S., Guptha, V., Kumari, G., Akhtar, S., Ekbal, A., Das, A., &

Chakraborty, T. (n.d.). *Fighting an Infodemic: COVID-19 Fake News Dataset*. <https://doi.org/10.1007/978-3>

Reis, J. C. S., Correia, A., Murai, F., Veloso, A., Benevenuto, F., & Cambria, E. (2019). Supervised Learning for Fake News Detection. *IEEE Intelligent Systems*, *34*(2), 76–81. <https://doi.org/10.1109/MIS.2019.2899143>

Sahoo, S. R., & Gupta, B. B. (2021). Multiple features based approach for automatic fake news detection on social networks using deep learning. *Applied Soft Computing*, *100*. <https://doi.org/10.1016/j.asoc.2020.106983>

United Nations. (2020). *During this coronavirus pandemic, 'fake news' is putting lives at risk: UNESCO | | 1UN news*. United Nations. Retrieved November 1, 2022, from <https://news.un.org/en/story/2020/04/1061592>

Wikimedia Foundation. (2023, January 23). *Misinformation*. Wikipedia. Retrieved January 28, 2023, from <https://en.wikipedia.org/wiki/Misinformation>

World Health Organization. (2020, June 29). *Listings of who's response to covid-19*. World Health Organization. Retrieved January 28, 2023, from <https://www.who.int/news/item/29-06-2020-covidtimeline>

World Health Organization. (2020, March 11). *Who director-general's opening remarks at the media briefing on COVID-19 - 11 march 2020*. World Health Organization. Retrieved January 28, 2023, from [https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-20](https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020)

J. Liu and M. Chen, "COVID-19 Fake News Detector," 2023 International Conference on Computing, Networking and Communications (ICNC), Honolulu, HI, USA, 2023, pp. 463-467, doi: 10.1109/ICNC57223.2023.10074216.

R. Malhotra, A. Mahur and Achint, "COVID-19 Fake News Detection System," 2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2022, pp. 428-433, doi: 10.1109/Confluence52989.2022.9734144.

Z. Tian and S. Baskiyar, "Fake News Detection using Machine Learning with Feature Selection," 2021 6th International Conference on Computing, Communication and Security (ICCCS), Las Vegas, NV, USA, 2021, pp. 1-6, doi: 10.1109/ICCCS51487.2021.9776346.

N. L. S. R. Krishna and M. Adimoolam, "Fake News Detection system using Decision Tree algorithm and compare textual property with Support Vector Machine algorithm," 2022 International Conference on Business Analytics for Technology and Security (ICBATS), Dubai, United Arab Emirates, 2022, pp. 1-6, doi: 10.1109/ICBATS54253.2022.9758999.