# TIME SERIES ANALYSIS USING CONVOLUTIONAL NEURAL NETWORKS (CNN) AND RECURRENT NEURAL NETWORKS (RNN) FOR MODELING THE FREQUENCY OF INFECTIOUS DISEASE EPIDEMIC NEWS

**(Thesis)**

**By**

**CLARISA SEPTIA DAMAYANTI**
**NPM. 2117031063**

**FACULTY OF MATHEMATICS AND NATURAL SCIENCES**
**UNIVERSITAS LAMPUNG**
**BANDAR LAMPUNG**
**2025**

# ABSTRACT

**TIME SERIES ANALYSIS USING CONVOLUTIONAL NEURAL NETWORKS (CNN) AND RECURRENT NEURAL NETWORKS (RNN) FOR MODELING THE FREQUENCY OF INFECTIOUS DISEASE EPIDEMIC NEWS**

By

**Clarisa Septia Damayanti**

The COVID-19 pandemic has highlighted the importance of leveraging online data as a tool for predicting future infectious disease trends. This study aims to compare the performance of two deep learning methods, namely Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN), in predicting the daily frequency of online news publications based on three sentiment classes: negative, neutral, and positive. The results show that CNN delivers the best performance, with an RMSE of 0.14 and MAPE of 27%, demonstrating its superiority in recognizing complex patterns in large datasets, especially for negative and neutral sentiment data. Meanwhile, RNN also yields reasonably good results, particularly for smaller datasets such as those with positive sentiment, although with slightly lower accuracy (RMSE of 0.17 and MAPE of 35%). These findings suggest that CNN is highly recommended for predictions on large-scale datasets, while RNN serves as a relevant alternative when data availability is limited, albeit with a slightly lower accuracy rate. Overall, deep learning models have proven effective in predicting the frequency of online news publications based on sentiment, supporting the use of online news as an alternative data source for monitoring public health issues.

**Keywords:** COVID-19, infectious diseases, online news, sentiment.

# ABSTRACT

## TIME SERIES ANALYSIS USING CONVOLUTIONAL NEURAL NETWORKS (CNN) AND RECURRENT NEURAL NETWORKS (RNN) FOR MODELING THE FREQUENCY OF INFECTIOUS DISEASE EPIDEMIC NEWS

**By**

**Clarisa Septia Damayanti**

Pandemi COVID-19 telah menyoroti pentingnya pemanfaatan data online sebagai alat untuk memprediksi tren penyakit menular di masa depan. Studi ini bertujuan untuk membandingkan kinerja dua metode deep learning, yaitu Convolutional Neural Network (CNN) dan Recurrent Neural Network (RNN), dalam memprediksi frekuensi harian publikasi berita online berdasarkan tiga kelas sentimen: negatif, netral, dan positif. Hasil penelitian menunjukkan bahwa CNN memberikan kinerja terbaik, dengan nilai RMSE sebesar 0,14 dan MAPE sebesar 27%, menunjukkan keunggulannya dalam mengenali pola kompleks pada dataset besar, terutama untuk data sentimen negatif dan netral. Sementara itu, RNN juga menghasilkan performa yang cukup baik, khususnya untuk dataset yang lebih kecil seperti data dengan sentimen positif, meskipun dengan tingkat akurasi yang sedikit lebih rendah (RMSE sebesar 0,17 dan MAPE sebesar 35%). Temuan ini menunjukkan bahwa CNN sangat direkomendasikan untuk prediksi pada dataset berskala besar, sementara RNN merupakan alternatif yang relevan ketika ketersediaan data terbatas, meskipun dengan tingkat akurasi yang sedikit lebih rendah. Secara keseluruhan, model deep learning terbukti efektif dalam memprediksi frekuensi publikasi berita online berdasarkan sentimen, sehingga mendukung penggunaan berita online sebagai sumber data alternatif untuk pemantauan isu kesehatan masyarakat.



**Kata kunci:** COVID-19, penyakit menular, berita online, sentimen.

# TIME SERIES ANALYSIS USING CONVOLUTIONAL NEURAL NETWORKS (CNN) AND RECURRENT NEURAL NETWORKS (RNN) FOR MODELING THE FREQUENCY OF INFECTIOUS DISEASE EPIDEMIC NEWS

**CLARISA SEPTIA DAMAYANTI**

**Thesis**

As One of the Requirements to Obtain the Degree
BACHELOR of MATHEMATICS

**Department of Mathematics**

**Faculty of Mathematics and Natural Sciences**

**FACULTY OF MATHEMATICS AND NATURAL SCIENCES**
**UNIVERSITAS LAMPUNG**
**BANDAR LAMPUNG**
**2025**

| | | |
|---|---|---|
| Thesis Title | : | **TIME SERIES ANALYSIS USING CONVOLUTIONAL NEURAL NETWORKS (CNN) AND RECURRENT NEURAL NETWORKS (RNN) FOR MODELING THE FREQUENCY OF INFECTIOUS DISEASE EPIDEMIC NEWS** |
| Student Name | : | *Clarisa Septia Damayanti* |
| Student ID | : | **2117031063** |
| Study Program | : | **Mathematics** |
| Faculty | : | **Faculty of Mathematics and Natural Sciences** |

**APPROVED**

1. Supervising Committee

**Dr. Khoirin Nisa, S.Si., M.Si.**
NIP 197407262000032001

**Dr. Purnomo Husnul Khotimah, M.T.**
NIP 198002062003121003

2. Head of the Department Mathematics

**Dr. Aang Nuryaman, S.Si., M.Si.**
NIP. 197403162005011001

**APPROVED**

1. Examination team

| | | | |
|---|---|---|---|
| Chair | : | **Dr. Khoirin Nisa, S.Si., M.Si.** | ........................ |
| Secretary | : | **Dr. Purnomo Husnul Khotimah, M.T.** | ........................ |
| Examiner Non-Supervisor | : | **Drs. Nusyirwan, M.Si.** | ........................ |

2. Dean of the Faculty of Mathematics and Natural Sciences

**Dr. Eng. Heri Satria, S.Si., M.Si.**
NIP. 197110012005011002

Date of Thesis Examination: **May 20, 2025**

## STUDENT THESIS STATEMENT

The undersigned:

| | | |
|---|---|---|
| Name | : | **Clarisa Septia Damayanti** |
| Student ID | : | **2117031063** |
| Department | : | **Mathematics** |
| Thesis Title | : | **TIME SERIES ANALYSIS USING CONVOLUTIONAL NEURAL NETWORKS (CNN) AND RECURRENT NEURAL NETWORKS (RNN) FOR MODELING THE FREQUENCY OF INFECTIOUS DISEASE EPIDEMIC NEWS** |

I hereby declare that this thesis is my own work. If it is later proven that this thesis is a copy or made by someone else, I am willing to accept sanctions according to the applicable academic regulations.

Bandar Lampung, 20 May 2025

Author,

Clarisa Septia Damayanti

## BIOGRAPHY

The author's name is Clarisa Septia Damayanti, born in Mekar Sari on September 13, 2003. She is the first child of two siblings, daughter of Mr. Jasni and Mrs. Yayuk Sri Sukaisih. She began her formal education at TK Radial Tutol, then continued to SD Negeri Mekar Sari from 2009 to 2015. She attended SMP Negeri 2 Pasir Sakti from 2015 to 2018 and completed her secondary education at SMA Negeri 1 Pasir Sakti, graduating in 2021. In 2021, she enrolled in the Mathematics Study Program, Faculty of Mathematics and Natural Sciences (FMIPA), University of Lampung.

During her studies, she actively participated in various organizational activities. She was a member of the Mathematics Student Association (HIMATIKA), involved in the cadre division at the department level, and was also active in the Student Executive Board (BEM) of FMIPA, University of Lampung, particularly in the Division of Research and Strategy (Dikasrat). Additionally, she volunteered at the Ruang Pangan organization, contributing as a Public Relations staff member.

The author also gained practical experience by participating in an internship program. She completed fieldwork internship (PKL) at the Central Statistics Agency (BPS) of Lampung Barat Regency. Subsequently, she undertook an internship at the National Research and Innovation Agency (BRIN), located at Jl. Cisitu Lama No. 37, Dago, Coblong District, Bandung City, West Java.

She further participated in a Certified Independent Study program via the RevoU platform, followed by the Certified Internship and Independent Study Program (MSIB) Batch 7 at The Farmhill, Semarang City, Central Java Province.

This thesis was prepared as one of the requirements to obtain a Bachelor of Mathematics degree at the University of Lampung.

# MOTTO

*”Indeed, with hardship [will be] ease.”*
(QS. Al-Insyirah: 4)

*”Allah does not burden a soul beyond that it can bear.”*
(QS. Al-Baqarah: 286)

*”Life is not about waiting for the storm to pass, but learning to dance in the rain.”*
(Anonymous)

*”I imagine my mother’s soul enduring to bring me into this world, so it is impossible that my existence has no meaning.”*

*”Place your trust solely in Allah, and He will open the way and provide for you at every step.”*

# ACKNOWLEDGMENT

With gratitude and thanks to Allah SWT for His blessings and guidance, enabling the completion of this thesis on time. With a sense of gratitude and happiness, I dedicate my thanks to:

**My Beloved Parents**

Thank you to my parents for all the sacrifices, motivation, prayers, and support throughout this journey. Thank you for teaching me valuable lessons about the true meaning of life, so that I can become a person who is beneficial to many.

**Supervisors and Examiners**

Thank you to my supervisors and examiners for their invaluable assistance, motivation, guidance, and knowledge.

**My Friends**

Thank you to all the wonderful people who have provided experiences, enthusiasm, motivation, prayers, and continuous support in every aspect.

**My Beloved Alma Mater**
Universitas Lampung

# PREFACE

Alhamdulillah, I express my praise and gratitude to Allah SWT for His abundant grace and blessings, allowing me to complete this thesis titled ”Time Series Analysis Using Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) for Modeling the Frequency of Infectious Disease Epidemic News” smoothly and on time. May blessings and peace be upon the Prophet Muhammad SAW. In the process of writing this thesis, many parties have contributed guidance, support, direction, motivation, and suggestions, enabling the completion of this work. Therefore, I would like to express my gratitude to:

1. Dr. Khoirin Nisa, S.Si., M.Si., as my first supervisor, who has consistently provided guidance, direction, motivation, suggestions, and support throughout the process of writing this thesis.

2. Dr. Purnomo Husnul Khotimah, M.T., as my second supervisor as well as my field supervisor during the internship, whose guidance and support have been invaluable both during the internship period and in the preparation of this thesis.

3. Drs. Nusyirwan, M.Si., as my examiner, who has provided constructive feedback, criticism, and valuable suggestions for the improvement of this thesis.

4. Dr. Aang Nuryaman, S.Si., M.Si. as the Head of the Mathematics Department, Faculty of Mathematics and Natural Sciences, Universitas Lampung.

5. To my beloved father, Mr. Jasni, and my dearest mother, Mrs. Yayuk Sri Sukaisih, who have always been a constant source of strength and encouragement throughout every step of my life. Thank you for your unwavering prayers, moral and financial support, and boundless, sincere love. Your sacrifices and dedication have been the greatest motivation throughout my academic journey. I also extend my heartfelt gratitude to my beloved sister, Anayya Nur

Fitriana, who has always brought joy and encouragement to every process I have gone through.

6. I would like to express my deepest gratitude to Yusuf Ihsan Al-Farizi for his invaluable assistance, support, and motivation throughout the research and thesis writing process.

7. Thank you to Anastasia, Yulina Putri, Kukuh Bayu, Al-Taqi, Amti, and Kak Fanny for their continuous support, help, and encouragement throughout this journey. Your presence and kindness have been instrumental in helping me complete this thesis.

8. Special thanks to my dear friends: Cantika, Lisa, Rani, Dinda, Mia, Gading, Fikri, and Dea, who have always stood by my side and supported me during my academic journey. Your friendship has made this experience more colorful and meaningful.

9. All lecturers, staff, and employees of the Department of Mathematics, Faculty of Mathematics and Natural Sciences, University of Lampung.

I hope this thesis can be beneficial for all of us. I realize that this work is far from perfect, and I welcome constructive criticism and suggestions to improve it further.

Bandar Lampung, 20 May 2025

Clarisa Septia Damayanti

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# CHAPTER I

# INTRODUCTION

## 1.1 Background of the Problem

The COVID-19 pandemic has underscored the importance of monitoring and predicting symptoms of infectious diseases to enable swift and effective responses. Real-time monitoring provides valuable insights into the severity and progression of diseases, allowing for the timely implementation of control and prevention measures (Lu et al., 2020). In this context, accurate predicting is essential to assess the potential for surges in cases that may pose public health threats (Fast et al., 2018). However, challenges in monitoring and predicting often arise due to the hierarchical structure of health information flow within existing systems.

As technology advances, digital epidemiology has emerged as a promising field that leverages data beyond traditional public health systems to study disease distribution. Digital epidemiology utilizes informal data sources to accelerate analysis without relying solely on primary health data (Salathé, 2018). One notable early initiative was Google Flu Trends (GFT), which used Google search query data to predict influenza trends in the United States (Olson et al., 2013). Additionally, online news has proven to be a valuable source of information for monitoring infectious diseases, including COVID-19 (Woo et al., 2016). Online news provides real-time, accessible information that can be analyzed by experts, making it an effective tool for tracking the progression of infectious diseases.

Previous studies have demonstrated the potential of using web data to monitor and forecast public health conditions (Khotimah et al., 2021). These studies analyzed the correlation between the frequency of online news coverage about COVID-19 and epidemiological data across various countries, revealing a strong correlation between online media frequency and the number of COVID-19 cases. This finding suggests that such data can be used in diagnosing and predicting infectious disease outbreaks.

Online news media not only play a crucial role in disseminating information to the public but also provide valuable insights for researchers and policymakers. With their ability to quickly and widely distribute information, these media cover a wide range of aspects, from disease development to government policies and preventive measures (Tang et al., 2018). In this regard, the frequency and sentiment of online news can reflect how public attention to infectious diseases evolves over time (Tsoy et al., 2021).

This study builds upon previous research (Khotimah et al., 2021) with the aim of utilizing Indonesian online news data related to COVID-19 to forecast the spread of infectious diseases. Two Deep Learning methods—Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN)—are compared to determine the most suitable approach for using online news frequency data in predicting disease trends. This study makes two primary contributions: first, it deepens the understanding of previous findings showing that the number of daily news articles correlates with the progression of infectious diseases and can be used for prediction; second, it provides technical insights into the use of Deep Learning models that are well-suited to the characteristics of the available data.

Given the global threats posed by infectious diseases such as COVID-19, avian flu, and Ebola—which require rapid and accurate predicting —this study compares the performance of Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) in classifying online news related to epidemics, to determine the most effective model. Sentiment analysis and news frequency are expected to identify patterns that may contribute to future spikes in disease cases (Song & Yoon, 2024). As a contribution to the development of more efficient and accurate predicting methods, this study aims to evaluate and compare the performance of CNN and RNN models in classifying online news related to epidemics. Identifying the most effective model is expected to serve as a foundation for selecting appropriate methods for analyzing and mitigating infectious diseases in the future, using online news data as a supporting information source (Adegoke et al., 2024).

## 1.2 Research Objectives

The objectives of this study are as follows:

1. To apply Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) models for predicting news related to epidemics.
2. To identify the most effective and efficient method for predicting news related to infectious disease outbreaks.
3. To compare the performance of Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) in predicting epidemic-related news.

## 1.3 Research Contributions

The contributions of this study are as follows:

1. To provide insights for decision-making in disease control and prevention through prediction.
2. To demonstrate the potential use of online news as an alternative data source for infectious disease analysis.
3. To establish a foundation for developing prediction models that utilize online news as input data.

# CHAPTER II

# LITERATURE REVIEW

## 2.1 Online News

Online news represents a significant development in the world of journalism that has emerged along with the rapid advancement of internet technology. Unlike traditional print media, online news allows information to be presented quickly and in real-time, providing access to audiences from various parts of the world to obtain the latest information. According to Mitchelstein and Boczkowski (2009), online news has changed the structure of news production, where journalists now more frequently rely on digital technology to disseminate news and face the demands to always provide news in a short time frame.

Online news also provides audiences with the opportunity to interact directly with news content through comments or sharing on social media, which influences the process of information dissemination and public participation in opinion formation (Allan, 2006). In some countries, such as Egypt, online news has brought significant changes in news consumption patterns, with audiences being more engaged and having broader access to various sources of information compared to traditional media (Mohammed and Gunter, 2013).

Online news articles provide rich data for various purposes, including natural language processing, sentiment analysis, and topic modeling. With the development of digital news platforms, the volume and variety of news articles have increased rapidly, making the collection and analysis of this data manually very challenging. Web scraping, as described by Rainey (2017), has become an effective solution for automating the process of extracting relevant data from online news articles. With the help of Python libraries such as BeautifulSoup and Scrapy, researchers can build web scrapers that navigate news sites, extract important information such as titles, dates, and article content, and store the data in a structured format for further analysis. This data can be used to explore trends, sentiments, and topics in the news, providing insights into public opinion, media bias, and the influence of news on society.

## 2.2 Epidemiology of Infectious Diseases

Epidemiology is the science focused on studying the patterns of distribution and spread of diseases within a population, as well as the factors influencing that spread (Porta et al., 2014). In the case of COVID-19, epidemiology plays a crucial role in understanding how the SARS-CoV-2 virus spreads and transmits among humans. COVID-19 itself is a disease caused by the SARS-CoV-2 virus, which was first detected in Wuhan, China in December 2019 (Zhou et al., 2020). Its spread occurs through direct contact with infected individuals, droplets from coughing or sneezing, and contaminated surfaces (World Health Organization, 2020).

## 2.3 Sentiment

Sentiment is the expression of emotions, opinions, or perceptions towards a specific topic, object, or situation that can be positive, negative, or neutral (Reshi et al., 2022). In data analysis, sentiment is often used to understand patterns of public opinion recorded through texts such as news, social media, or product reviews. Sentiment analysis becomes an important tool in assessing public reactions to an event or policy, especially in the context of public health and global crises such as the COVID-19 pandemic (Pristiyono et al., 2021).

Sentiment is used to identify the direction of opinions or information conveyed by news related to the COVID-19 pandemic. Sentiment is divided into three main categories:

a. **Negative Sentiment**: News that reports problems or worsening situations, such as *"Surge in Cases Threatens Hospitals"*. This news reflects concerns or negative impacts of the pandemic.

b. **Neutral Sentiment**: News that only provides factual information without positive or negative opinions, such as *"COVID-19 Health Tips"*.

c. **Positive Sentiment**: News that shows good developments or positive news related to the pandemic, such as *"Patients Show Negative Test Results"*. This category indicates optimism regarding the pandemic situation.

### 2.4 Time Series

Time series analysis is an important method in data processing used in various fields of research to understand patterns and trends that develop over time. A time series consists of a set of data measured periodically at specific time intervals, and it has proven to be very useful in various disciplines for forecasting future data movements (Hyndman and Athanasopoulos, 2018).

As an analytical tool, time series allows for the identification of historical patterns and long-term dynamics that can support data-driven decision-making. According to Hyndman and Athanasopoulos (2018), time series-based forecasting is a crucial element in fields such as economics, finance, and environmental science, as it provides the ability to anticipate changing conditions and respond to predicted developments.

In its development, researchers have developed various approaches to leverage time series analysis to predict future data based on historical information. By using this method, time series is not only used to forecast subsequent data values but also to analyze trend changes, seasonal patterns, and patterns that are not always directly visible in raw data (Ma et al., 2020). Recent studies have also explored more complex techniques for analyzing time series that contain difficult-to-detect patterns, especially when the data exhibits non-linear characteristics that require specialized methods for effective processing (Salman et al., 2024).

With advancements in technology, this research has evolved by utilizing more sophisticated computational techniques to enhance prediction accuracy. The methods used in time series analysis have also become increasingly diverse, encompassing both traditional approaches and more modern machine learning-based methods that can capture complex patterns and provide reliable results for various current data analysis needs (Hamilton, 2020).

## 2.5 Preprocessing

The data preprocessing process is a crucial stage in preparing the collected data before it is used in the model training process. This stage is essential to ensure that the available data can be optimally used in model training, resulting in good and accurate performance. The preprocessing process helps to clean, format, and adjust the data to better fit the analysis or model requirements that will be applied (Kotu dan Deshpande, 2018).

### 2.5.1 *Data Cleaning*

Data cleaning is a very important stage in data analysis to prepare the data for further analysis. This process includes identifying and correcting issues found in raw data, such as missing data, duplicates, or inconsistencies (Kotu dan Deshpande, 2018). Some common steps taken in data cleaning include:

1. Handling Missing Values: Missing data in time series can introduce bias in analysis. A common method to address this issue is *forward fill*, which replaces missing values with the previous values to maintain data continuity without altering trend patterns.

2. Removing Irrelevant Columns: Irrelevant columns or those with excessive missing values can be removed to reduce data dimensionality, speed up processing, and improve analysis accuracy without losing essential information.

3. Grouping Online News Sentiments: Online news can be classified based on sentiment (positive, neutral, negative) using lexicon-based analysis or machine learning models. This step is crucial for understanding public opinion trends on specific issues through digital media.

4. Data Filtering for Time Series Analysis: Data filtering is applied to eliminate anomalies or extreme values that may affect the analysis. Additionally, time interval adjustments are performed to ensure more uniform data, making it suitable for modeling.

5. Calculating News Sentiment Frequency: Measuring the daily number of news articles based on sentiment helps analyze trends and the media's impact on public opinion. This data is commonly used in studies examining the relationship between news coverage and issue developments.

According to Kotu dan Deshpande, (2018), data cleaning is an indispensable step in data analysis, as the quality of the analysis results heavily depends on the quality of the data used. Without proper data cleaning, the results of the analysis or predictive models may lead to incorrect conclusions.

### 2.5.2 Data Normalization

Data normalization or data scaling is a process that transforms the original data into a more manageable range of values, allowing the analysis model or prediction to work optimally with minimal error. One technique used is MinMaxScaler, which transforms the data into the range [0,1]. This technique is formulated as follows (Kotu dan Deshpande 2018):

$$x' = \frac{x - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}}$$

With the following explanation:

$x'$: The normalized online news frequency data.

$x$: The actual online news frequency data.

$X_{\text{min}}$: The minimum value of the online news frequency data.

$X_{\text{max}}$: The maximum value of the online news frequency data.

### 2.5.3 Data Splitting for Training and Testing

Data splitting is a crucial step in data analysis that involves dividing the dataset into two parts, namely training data and testing data. Training data is used to train the model, while testing data is used to evaluate the accuracy or performance of the trained model. The primary goal of this data splitting is to obtain a model that can provide the best possible results (Cahyadi et al., 2020).

## 2.6 *Deep Learning*

Deep learning is a branch of machine learning that uses artificial neural networks to build hierarchical representations of data through multiple layers of processing. As stated by Ilahiyah dan Nilogiri (2018), deep learning has proven to be effective in various modern data applications. Furthermore, Openg et al. (2022) adds that this approach is particularly useful for capturing complex relationships and temporal dependencies, which are highly relevant in time series analysis. This approach helps capture complex and dynamic patterns in sequential data (Chen et al., 2018).

In research conducted by Zhang (Zhang, 2019), it is explained that various deep learning algorithms have the advantage of extracting features from historical data that can improve prediction accuracy. Zhang emphasizes that deep learning's ability to handle large and complex data volumes makes it a valuable tool in time series analysis, especially in the era of rapidly evolving information.

Moreover, Wang et al. (Wang et al., 2019) demonstrate that deep learning is not only capable of capturing complex patterns in time series data but also can improve prediction accuracy through the use of deeper and more complex network architectures. This research provides evidence that integrating deep learning techniques into time series analysis yields better results compared to traditional methods, thus making a significant contribution to understanding data dynamics in various applications.

Overall, the application of deep learning in time series analysis opens up new opportunities for further research, with the potential to model and predict future data more effectively. Therefore, the use of these advanced techniques is expected to provide deeper insights into patterns and trends in data, supporting more accurate data-driven decision-making.

## 2.7 Convolutional Neural Network (CNN)

Convolutional Neural Networks (CNN) are an extension of Multilayer Perceptron (MLP) designed to process two-dimensional data. As a type of Deep Neural Network, CNN is characterized by its high depth and is widely applied to image data.

In image classification tasks, MLP is less suitable because of its inability to preserve spatial information from image data, treating each pixel as an independent feature, which leads to less optimal results.

Convolutional Neural Networks (CNN) can be applied to time series or sequential data, such as audio, signal processing, or financial time series data, to extract features from sequential data. In this context, 1D convolutional layers are used to process sequential data, scanning the input sequence one by one, and followed by pooling layers to downsample the feature maps, preserving important information. The output is then flattened into a 1D feature vector and fed into dense layers for classification, regression, or prediction tasks. CNN is robust to noise and irregularities in time series data, can capture non-linear relationships and patterns, and can handle input sequences of varying lengths, making it suitable for time series data with varying lengths. CNN has various applications, including anomaly detection, time series classification, and prediction, such as identifying unusual patterns in sensor readings, classifying audio signals or stock prices, or predicting future stock prices or energy consumption.

### 2.7.1 CNN Concept

The main concept of CNN is the application of convolutional layers to extract features from input data. Convolutional layers consist of filters applied to small areas of the input data (e.g., a group of pixels in an image or a sequence of values in a time series). These filters are responsible for detecting patterns, such as edges in images or changes in trends in time series data.

Convolutional layers are followed by pooling layers that reduce the dimensionality of the feature maps produced, allowing for reduced computational complexity and increased efficiency. Pooling is performed by taking the maximum (max pooling) or average (average pooling) value from a specific area of the data.

After several convolutional and pooling layers, the output is flattened into a 1D vector and then passed through fully connected (dense) layers, where the final prediction is made. The network is trained using the backpropagation algorithm to optimize the weights on the filters and dense layers, effectively learning relevant patterns from the input data.

### 2.7.2 CNN Architecture

The CNN architecture consists of several types of layers that function to extract important features and make predictions. Some of the main layers in the CNN architecture are:

1. Convolution Layer Convolutional layers are the core of CNN, where filters are applied to the input to extract local features from the data. Each filter scans small parts of the input (e.g., a group of pixels or a sequence of values) and produces a feature map that highlights important patterns. Multiple filters can be applied to detect various features in the data, such as edges, textures, or changes in trends in time series data.
2. Subsampling Layer (Pooling Layer) Subsampling layers, also known as pooling layers, are used to reduce the dimensionality of the feature maps produced by convolutional layers. Pooling is performed by taking the maximum (max pooling) or average (average pooling) value from a specific area of the feature map. Reducing dimensionality helps reduce computational complexity and prevent overfitting, while preserving important information from the data.
3. Fully Connected Layer After several convolutional and pooling layers, the output is flattened into a 1D feature vector and fed into fully connected (dense) layers. Fully connected layers function like MLP, where each neuron is connected to all neurons in the previous layer. In this layer, the final decision is made, such as classification or regression. Fully connected layers are responsible for integrating all the features extracted earlier to produce the final prediction.

Overall, the CNN architecture provides the ability to capture complex and hierarchical patterns in data, whether in the context of images, signals, or time series, making it a highly flexible and powerful tool for various applications.

## 2.8 Recurrent Neural Network (RNN)

Recurrent Neural Networks (RNN) are a type of Artificial Neural Networks (ANN) that can identify hidden patterns in data for various applications such as speech recognition, natural language processing, and time series forecasting (Chen et al., 2018). RNN is highly effective in handling problems that involve sequences because of its ability to process input information along with historical information through recurrent connections.

This neural network is called recurrent because the output from the previous hidden layer is used again as input in the next process (Hewamalage et al., 2021). A characteristic of RNN is its ability to not only depend on the input but also on the input received previously, so that each input is interrelated and can provide additional information to the hidden layer (Hewamalage et al., 2021).

According to Hewamalage et al. (2021), RNN consists of input layers, hidden layers, and output layers. The RNN model has an architecture with information flowing from the input layer to the hidden layer, as well as the process of storing information from the previous hidden layer for use in the current hidden layer. This RNN model architecture is designed to handle information flow recurrently.

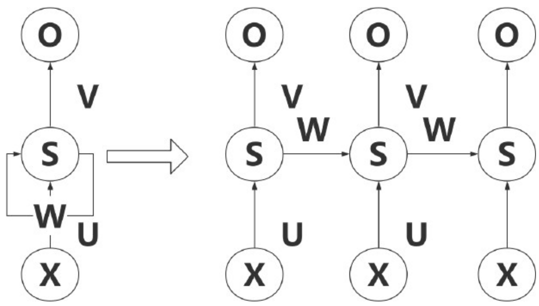


**Figure 2.1 RNN Architecture**

As shown in Figure 2.1, the mapping of $S_t$ and $O_t$ can be represented as follows:

$$S_t = f(U \times X_t + W \times S_{t-1}) \tag{2.3}$$

$$O_t = g(V \times S_t) \tag{2.4}$$

Where:

$S_t$: The network memory at time t

$U, V,$ and $W$: Weight matrices in each layer

$X_t$ and $O_t$: Input and output at time t

$f(\cdot)$ and $g(\cdot)$: Non-linear functions

### 2.8.1 Activation Functions in RNN

Activation functions in neural networks are used to activate or not activate neurons (Qamar & Zardari, 2023). The activation functions commonly used are the sigmoid activation function and the hyperbolic tangent activation function.

### 2.8.2 Sigmoid Activation Function

The sigmoid activation function is a non-linear function that has a range of values between 0 and 1. The sigmoid function is shown in the following equation:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{2.5}$$

Figure 2.2 shows the graph of the sigmoid activation function.

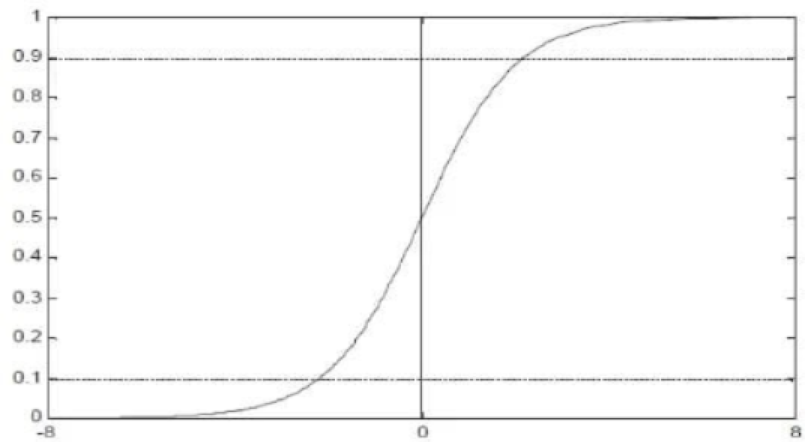


**Figure 2.2 Sigmoid Activation Function**

### 2.8.3 Hyperbolic Tangent Activation Function

The Hyperbolic Tangent Activation Function, or commonly referred to as the tanh function, is an alternative to the sigmoid function. The tanh function has a range of values between -1 and 1. According to (Nwankpa et al., 2018), the tanh function is formulated as follows:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{2.6}$$

Figure 2.3 shows the graph of the hyperbolic tangent activation function.

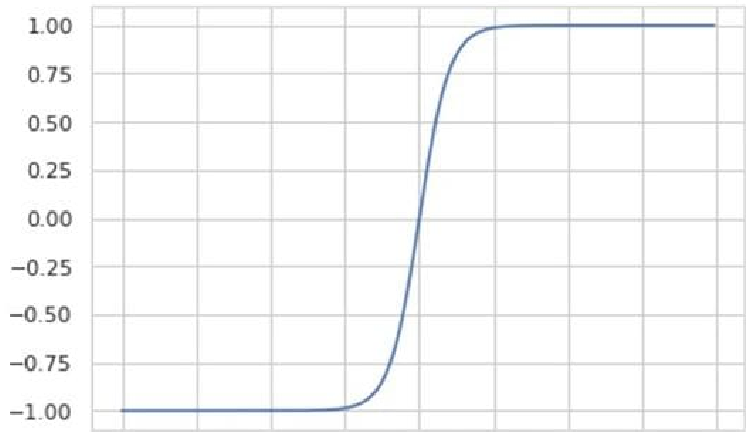


**Figure 2.3 Hyperbolic Tangent Activation Function**

The sigmoid and tanh activation functions are two of the many activation functions used in neural networks. According to (Goodfellow et al. ,2016), the choice of the appropriate activation function is crucial for the performance of the model. (Haykin ,2009) also emphasizes that different activation functions can affect the convergence and speed of neural network training. In addition, (Bishop Nasrabadi,2006) explains that a good understanding of activation functions can help in designing more effective network architectures.

## 2.9 Parameter Initialization

To produce an optimal model, hyperparameter tuning is crucial, especially in terms of the number of neurons and dropout. Dropout serves to reduce overfitting and speed up learning in neural networks (Faishol et al., 2020). Overfitting occurs when the model is very accurate on the training data but fails to predict new data well, thus reducing its ability to generalize. Dropout works by temporarily disabling some neurons in the hidden layer to reduce excessive dependence on specific paths.

In addition, parameters such as epochs and batch size also play a crucial role. According to Putra et al. (2022), an epoch is the number of times the entire dataset is repeated during network training. One epoch includes all the data being processed, but too large an epoch can slow down training. To address this, the data is broken down into small batches, known as batch size. Batch size refers to the number of samples processed before the network weights are updated, speeding up and increasing the efficiency of training.

## 2.10 Denormalization

Denormalization is a process used to transform the predicted results that have been normalized back to their original data form. This step is necessary to compare the predicted data with the actual data to evaluate the performance of the built model. If the data was previously normalized in the range [0,1], then denormalization is performed using the following equation (Wiranda dan Sadikin,2019):

$$y_t = y'(X_{max} - X_{min}) + X_{min} \tag{2.13}$$

Where:

$y_t$: The value of the predicted online news frequency data that has been denormalized.

$y'$: The predicted online news frequency data that has been normalized.

$X_{min}$: The minimum value of the online news frequency data.

$X_{max}$: The maximum value of the online news frequency data.

## 2.11 Model Evaluation

Model evaluation is the process of testing the accuracy of the performance of the model formed in predicting data that has been tested previously to determine the performance of the trained model. According to Wiranda dan Sadikin (2019), to evaluate the model, Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) can be used.

### 2.11.1 Root Mean Square Error (RMSE)

Root Mean Square Error (RMSE) is a method used to evaluate a model by calculating the difference between actual data and predicted data. The RMSE value is calculated by subtracting the actual data and prediction, squaring the difference, calculating the average, and then taking the square root. The RMSE formula is as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(Y_i - \hat{Y}_i)^2} \tag{2.14}$$

Where:

$Y_i$: The online news frequency data at time t.

$\hat{Y}_i$: The predicted online news frequency data at time t.

$n$: The number of periods of online news frequency data.

### 2.11.2 Mean Absolute Percentage Error (MAPE)

Mean Absolute Percentage Error (MAPE) measures the percentage error between the actual data value and its predicted value. MAPE is formulated with the following equation:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{Y_i - \hat{Y}_i}{Y_i}\right| \times 100 \tag{2.15}$$

Where:

$Y_i$: The online news frequency data at time t.

$\hat{Y}_i$: The predicted online news frequency data at time t.

$n$: The number of periods of online news frequency data.

If the MAPE value of the model is smaller, the prediction value becomes closer to the actual value. The model with the smallest MAPE value is considered the best model for forecasting online news frequency. The MAPE value categories for measuring the model's performance are shown in Table 2.1 below (Hayuningtyas, 2017).

As shown in Table 2.1, the MAPE value categories are as follows:

**Table 2.1 MAPE Value Range Categories**

| MAPE Range | Category |
|---|---|
| $<$10% | Very good forecasting model |
| 10%-20% | Good forecasting model |
| 20%-50% | Adequate forecasting model |
| $>$50% | Poor forecasting model |

# CHAPTER III

# RESEARCH METHOD

## 3.1 Research Time and Place

This research was conducted in the second semester of the 2023/2024 academic year, spanning 6 months from February to July at the Data Science and Information Center, National Research and Innovation Agency (BRIN), located at Sangkuriang Street, Dago, Bandung, West Java, 40135, Indonesia. This center is part of BRIN and focuses on the development of data science and information management for various research and innovation needs in Indonesia.

Table 3.1 presents the research activity timeline carried out during the specified period.

**Table 3.1 Research Activity Timeline**

| **Activity** | **February** | **March** | **April** | **May** | **June** | **July** |
|---|---|---|---|---|---|---|
| Literature Review | ✓ | ✓ | | | | |
| Data Collection | | ✓ | ✓ | | | |
| Data Preprocessing | | | ✓ | ✓ | | |
| Model Design | | | | ✓ | ✓ | |
| Model Testing | | | | | ✓ | ✓ |
| Model Comparison | | | | | | ✓ |
| Thesis Writing | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

## 3.2 InaCOVED Dataset

The *Indonesian Corpus for COVID-19 Event Detection (InaCOVED)* is a collection of online news articles gathered from seven major news portals in Indonesia: Tirto, Tempo, Republika, Merdeka, Kompas, Detik, and Antara. This dataset was compiled by the Information Retrieval Research Group at the BRIN Informatics Research Center and has been labeled for research purposes. Designed to detect important events related to the COVID-19 pandemic in Indonesia, *InaCOVED* covers various relevant topics, including the development of COVID-19 cases, health policies, government responses, and public reactions to the pandemic (Khotimah 2023).

In this research, *InaCOVED* is used to analyze online news related to COVID-19. The dataset includes a total of 16,839 articles accumulated during the data collection period, from January 26 to May 24, 2020. This data provides important insights into the development of COVID-19 cases in Indonesia, the health policies implemented during the pandemic, as well as the reactions of the government and the public to the situation.

The following Table (3.2) shows the distribution of articles from each news portal that contributed to the InaCOVED dataset:

**Table 3.2 Article Distribution from News Portals**

| Portal | Number of Articles |
|---|---|
| Tirto | 127 |
| Tempo | 2,438 |
| Merdeka | 1,109 |
| Republika | 2,560 |
| Kompas | 3,073 |
| Detik | 4,280 |
| Antara | 3,254 |
| **Total** | **16,839** |

The *InaCOVED* dataset consists of seven columns: `_id`, `portal`, `published_at`, `title`, `title_clean`, `event`, and `sentiment`.

The following Table 4.1 presents a detailed description of each column in the dataset:

**Table 3.3 Dataset Columns Description**

| Column | Description |
|---|---|
| `_id` | Unique identification number generated by MongoDB. Example: ObjectId ("6130884e9067153d079ab188"). |
| `portal` | Name of the associated news portal. Example: Tirto, Tempo, Republika, Merdeka, Kompas, Detik, and Antara. |
| `published_at_iso` | Date and time of news publication in ISO format. Example: ISODate("2021-07-24T21:35:06.1552"). |
| `title` | News title. Example: 1,591 Patients Recover from Covid-19, DKI Jakarta Has the Most. |
| `title (without ., ?, ", ')` | News title without punctuation marks like periods (.), commas (,), question marks (?), double quotes ("), and single quotes ('). Example: 1591 Patients Recover from Covid-19 DKI Jakarta Has the Most. |
| `event` | News label as event/non-event. Example: 1: event, 0: non-event. |
| `sentiment` | Sentiment label of the news. Example: -1: negative, 0: neutral, 1: positive. |

## 3.3 Research Methods

This research aims to identify the best method between **Convolutional Neural Network (CNN)** and **Recurrent Neural Network (RNN)** for predicting the frequency of online news articles classified into sentiment categories: positive, neutral, and negative, using Python on Google Colab. The research procedures are as follows:

### 3.3.1 Research Procedures for the Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) Methods

1. Preparing the INACOVED dataset, which includes COVID-19 news that has been classified into three sentiment categories: positive, neutral, and negative.
2. Performing data preprocessing:
   a. Handling missing values using the *forward fill* technique.
   b. Removing irrelevant columns.
   c. Grouping the data based on sentiment (positive, neutral, negative).
   d. Preparing the data for time series analysis using `data filtering`.
   e. Calculating the number of news articles for each sentiment category (positive, neutral, and negative) on a daily basis.
   f. Normalizing the data using scaling methods to ensure well-distributed data so that the model can handle data spikes.
3. Splitting the data into 80% for training and 20% for testing.
4. Determining CNN and RNN model parameters such as the number of filters, kernel size, dropout rate, batch size, and number of epochs.
5. Building the CNN and RNN model and performing hyperparameter tuning to optimize parameters such as the number of filters, kernel size, dropout rate, number of hidden units, batch size, and number of epochs.
6. Making predictions using the trained CNN and RNN model.
7. Denormalizing the prediction results to allow comparison with the actual data.
8. Evaluating the performance of the CNN and RNN model using RMSE (Root Mean Squared Error) and MAPE (Mean Absolute Percentage Error) metrics.
9. Plotting the predicted and actual data for visual comparison.

### 3.3.2 Model Analysis and Comparison

After evaluating the performance of each model, this study conducts an analysis and interpretation of the results to gain insights into the sentiment data. The analysis focuses on identifying the strengths and weaknesses of each model, as well as observing trends and patterns within the data. The performance of the models is compared using RMSE (Root Mean Squared Error) and MAPE (Mean Absolute Percentage Error) metrics to determine the most accurate and effective model for prediction.

### 3.3.3 Trend Analysis

This study analyzes trends and patterns in the sentiment data to identify seasonal or temporal dependencies. This analysis helps to understand the dynamics of sentiment changes over time and relate them to specific events or factors that may influence the data.

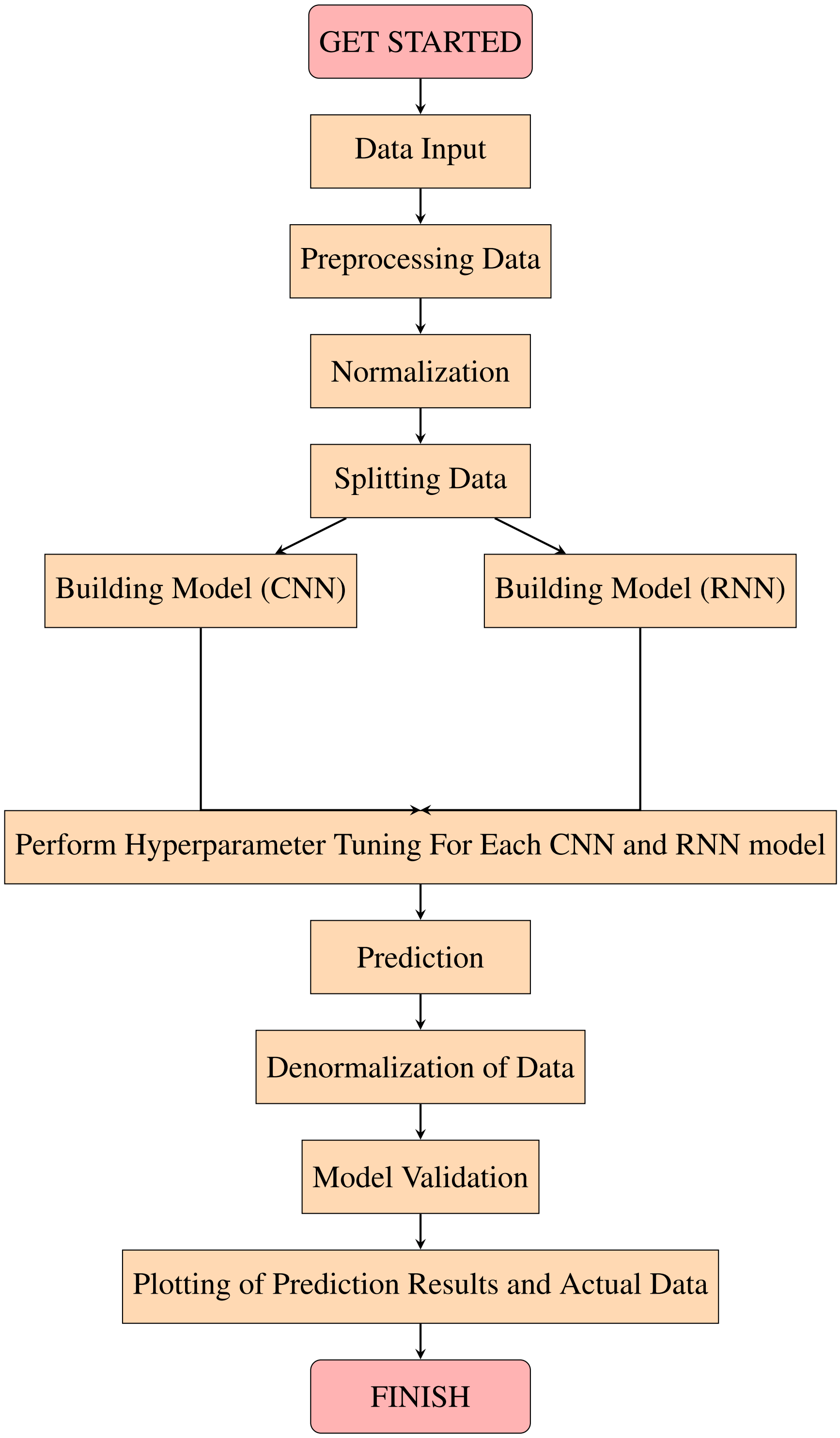


**Figure 3.1 Research Flowchart**

# CHAPTER V

# CONCLUSION AND RECOMMENDATIONS

## 5.1 Conclusion

This study compares the performance of CNN and RNN models in predicting the daily frequency of online news articles based on sentiment classes related to COVID-19. The results show that CNN delivers the best performance with an RMSE of 0.14 and MAPE of 27%, demonstrating its superiority in recognizing complex patterns in large datasets, such as negative and neutral sentiment data. Meanwhile, RNN also shows fairly good results, particularly with smaller datasets like positive sentiment, with an RMSE of 0.17 and MAPE of 35%. Although its accuracy is slightly lower than CNN, RNN remains a relevant alternative when data is limited. These findings indicate that both CNN and RNN have their respective strengths. CNN is recommended for predictions on large-scale datasets, while RNN can serve as a good alternative when available data is limited, albeit with a slightly lower accuracy rate. Overall, deep learning models have proven effective in predicting the frequency of online news based on sentiment, supporting the use of digital news as an alternative data source for monitoring public health issues.

## 5.2 Suggestions

Based on the results of this study, it is recommended that future researchers compare the performance of CNN and RNN with traditional statistical models such as ARIMA or linear regression, and explore other machine learning models like Decision Trees, SVM, or XGBoost to predict the frequency of news publications related to COVID-19 or other infectious diseases.

# REFERENCES


Adegoke, B. O., Odugbose, T., & Adeyemi, C. (2024). Data analytics for predicting disease outbreaks: A review of models and tools. *International Journal of Life Science Research Updates*, 2(2), 1-9. https://doi.org/10.5281/zenodo.1234567

Allan, S. (2006). *Online News: Journalism and the Internet*. McGraw-Hill Education (UK).

Bishop, C. M., Nasrabadi, N. M. (2006). *Pattern Recognition and Machine Learning* (Vol. 4, No. 4, p. 738). New York: Springer.

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (2015). *Time Series Analysis: Forecasting and Control* (5th ed.). John Wiley & Sons.

Cahyadi, R., Damayanti, A., & Aryadani, D. (2020). Recurrent Neural Network (RNN) dengan Long Short Term Memory (LSTM) untuk Analisis Sentimen Data Instagram. *Jurnal Informatika dan Komputer (JIKO)*, 5(1), 1–9.

Carpentier, N., et al. (2018). *The Impact of Online News Consumption on Political Participation: Evidence from the 2016 US Presidential Election*. Journal of Broadcasting & Electronic Media, 62(4), 636–654.

Charles-Smith, L. E., Reynolds, T. L., Cameron, M. A., Conway, M., Lau, E. H., Olsen, J. M., ... & Corley, C. D. (2015). Using social media for actionable disease surveillance and outbreak management: a systematic literature review. *PloS One*, 10(10), e0139701. https://doi.org/10.1371/journal.pone.0139701

Chen, Y., Cheng, Q., Cheng, Y., Yang, H., Yu, H. (2018). Applications of recurrent neural networks in environmental factor forecasting: a review. *Neural Computation*, 30(11), 2855–2881. https://doi.org/10.1162/neco_a_01156

Chen, Y., Zhang, Y., and Chen, L. (2018). *Deep Learning for Time Series Forecasting: A Survey*. IEEE Transactions on Neural Networks and Learning Systems, 29(10), 4524–4536.

Chung, J., Gulcehre, C., Cho, K., and Bengio, Y. (2014). *Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling*. arXiv preprint arXiv:1412.3555.

Deshpande, R. R. (2012). On the rainfall time series prediction using multilayer perceptron artificial neural network. *International Journal of Emerging Technology and Advanced Engineering*, 2(1), 2250-459.

Egan, S., Fedorko, W., Lister, A., Pearkes, J., Gay, C. (2017). *Long Short-Term Memory (LSTM) networks with jet constituents for boosted top tagging at the LHC*. arXiv preprint arXiv:1711.09059.

Faishol, M. A., Endroyono, Irfansyah, A. N. (2020 ). Prediction of Urban Air Pollution in Surabaya Using Recurrent Neural Network–Long Short Term Memory. *Jurnal Ilmiah Teknologi Informasi*, 18(2), 102-114.

Fast, S. M., Kim, L., Cohn, E. L., Mekaru, S. R., Brownstein, J. S., Markuzon, N. (2018). Predicting social response to infectious disease outbreaks from internet-based news streams. *Annals of Operations Research*, 263, 551–564.

Gers, F. A., Schmidhuber, J., Cummins, F. (2000). *Learning to forget: Continual prediction with LSTM*. Neural Computation, 12(10), 2451-2471. https://doi.org/10.1162/089976600300015015

Goodfellow, I., Bengio, Y., Courville, A. (2016). *Deep Learning*. MIT Press.

Hamilton, J. D. (2020). *Time series analysis* (2nd ed.). Princeton University Press.

Haykin, S. (2009). *Neural Networks and Learning Machines* (3rd ed.). Pearson Education India.

Hayuningtyas, R. Y. (2017). Peramalan Persediaan Barang Menggunakan Metode Weighted Moving Average dan Metode Double Exponential Smoothing. *Jurnal PILAR Nusa Mandiri*, 13(2), 217–219.

Hewamalage, H., Bergmeir, C., Bandara, K. (2021). Recurrent neural networks for time series forecasting: Current status and future directions. *International Journal of Forecasting*, 37(1), 388–427. https://doi.org/10.1016/j.ijforecast.2020.08.003

Hochreiter, S., Schmidhuber, J. (1997). *Long short-term memory*. Neural Computation, 9(8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hyndman, R. J., and Athanasopoulos, G. (2018). *Forecasting: Principles and Practice*. OTexts.Retrieved from `https://otexts.com/fpp2/`

Ilahiyah, dan Nilogiri, (2018). *Deep Learning: A Review*. Journal of Intelligent Information Systems, 51(2), 231–244.

Keshavamurthy, R., Dixon, S., Pazdernik, K. T., & Charles, L. E. (2022). Predicting infectious disease for biopreparedness and response: A systematic review of machine learning and deep learning approaches. *One Health*, 15, 100439. https://doi.org/10.1016/j.onehlt.2022.100439

Khotimah, P., Latifah, A., Syafarina, I., Shabrina, A., Nugraheni, E., & Rozie, A. (2021). COVID-19 vs media: Time series data analysis towards media and COVID-19 epidemic data. In *Proceedings of the 2021 International Conference on Computer, Control, Informatics and Its Applications* (pp. 156–160).

Khotimah, P. H. (2023). InaCOVED (Indonesian Corpus for COVID-19 Event Detection) on Online News. [Online]. Available: `https://hdl.handle.net/20.500.12690/RIN/8K14KW`.

Khotimah, P. H., Arisal, A., Rozie, A. F., Nugraheni, E., Riswantini, D., Suwarningsih, W., Munandar, D., & Purwarianti, A. (2023). Monitoring Indonesian online news for COVID-19 event detection using deep learning. *International Journal of Electrical & Computer Engineering*, 13(1).

Kotu, V., & Deshpande, B. (2018). *Data science: concepts and practice*. Morgan Kaufmann.

Krawczyk, K., Chelkowski, T., Laydon, D. J., Mishra, S., Xifara, D., Gibert, B., Flaxman, S., Mellan, T., Schwämmle, V., & Röttger, R. et al. (2021). Quantifying online news media coverage of the COVID-19 pandemic: Text mining study and resource. *Journal of Medical Internet Research*, 23(6), e28253.

Lai, T. L., and Xing, H. (2017). Machine Learning for Time Series Forecasting. *Journal of Machine Learning Research, 18(1), 1–35.*

Liu, S., Chen, Y., He, S., Liu, K., Zhao, J. (2016). Leveraging FrameNet to improve automatic event detection. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 2134–2143).

Little, R. J., & Rubin, D. B. (2019). *Statistical analysis with missing data* (Vol. 793). John Wiley & Sons.

Lu, H., Stratton, C. W., Tang, Y.-W. (2020). The Wuhan SARS-CoV-2—what's next for China. *Journal of Medical Virology*, 92(6), 546.

Ma, T., Antoniou, C., Toledo, T. (2020). Hybrid machine learning algorithm and statistical time series model for network-wide traffic forecast. *Transportation Research Part C: Emerging Technologies*, 111, 352-372. `https://doi.org/10.1016/j.trc.2019.01.011`

Mitchelstein, E., and Boczkowski, P. J. (2009). Between Tradition and Change: A Review of Recent Research on Online News Production. *Journalism, 10(5), 562–586.*

Mohammed, H., and Gunter, B. (2013). *News Consumption and News Agendas in Egypt*. In *News Media in the Arab World: A Study of 10 Arab and Muslim Countries*, 83–105.

Nwankpa, C., Ijomah, W., Gachagan, A., Marshall, S. (2018). Activation functions: Comparison of trends in practice and research for deep learning. *arXiv preprint arXiv:1811.03378*.

Olson, D. R., Konty, K. J., Paladini, M., Viboud, C., Simonsen, L. (2013). Reassessing Google Flu Trends data for detection of seasonal and pandemic influenza: A comparative epidemiological study at three geographic scales. *PLoS Computational Biology*, 9(10), e1003256.

Openg, R. K., et al. (2022).Deep Learning for Time Series Forecasting: A Review. *Journal of Forecasting, 41(1), 1–23.*

Paul, M. J., Dredze, M., Broniatowski, D. (2014). Twitter improves influenza forecasting. *PLoS Currents*, 6.

Porta, M. S., Greenland, S., Hernán, M., dos Santos Silva, I., Last, J. M. (Eds.). (2014). *A dictionary of epidemiology*. Oxford University Press.

Pristiyono, Ritonga, M., Ihsan, M. A. A ., Anjar, A., Rambe, F. H. (2021). Sentiment analysis of COVID-19 vaccine in Indonesia using Naïve Bayes Algorithm. *IOP Conference Series: Materials Science and Engineering*, 1088(1), 012045. IOP Publishing.

Putra, R. B. R., & Hendry. (2022). Multivariate Time Series Forecasting pada Penjualan Barang Retail dengan Recurrent Neural Network. *Jurnal Inovtek Polbeng-Seri Informatika*, 7(1), 71-82.

Rainey, J. (2017). Web Scraping with Python: Collecting Data from the Modern Web. *" O'Reilly Media, Inc."*.

Rana, A., Rawat, A. S., Bijalwan, A., Bahuguna, H. (2018, August). Application of multi layer (perceptron) artificial neural network in the diagnosis system: a systematic review. In *2018 International conference on research in intelligent and computing in engineering (RICE)* (pp. 1-6). IEEE.

Reshi, A. A., Rustam, F., Aljedaani, W., Shafi, S., Alhossan, A., Alrabiah, Z., Ashraf, I. (2022). Covid-19 vaccination-related sentiments analysis: a case study using worldwide twitter dataset. *Healthcare*, 10(3), 411. MDPI.

Salathé, M. (2018). Digital epidemiology: what is it, and where is it going? *Life Sciences, Society and Policy*, 14(1), 1.

Salman, D., Direkoglu, C., Kusaf, M., Fahrioglu, M. (2024). Hybrid deep learning models for time series forecasting of solar power. *Neural Computing and Applications*, 1-18. https://doi.org/10.1007/s00500-023-06881-4

Smagulova, K., James, A. P. (2019). A survey on LSTM memristive neural network architectures and applications. *The European Physical Journal Special Topics, 228(10), 2313-2324* https://doi.org/10.1140/epjst/e2019-900045-0

Song, Y., & Yoon, B. (2024). Prediction of infectious diseases using sentiment analysis on social media data. *PloS One*, 19(9), e0309842. https://doi.org/10.1371/journal.pone.0309842

Tang, L., Bie, B., Park, S. E., & Zhi, D. (2018). Social media and outbreaks of emerging infectious diseases: A systematic review of literature. *American Journal of Infection Control*, 46(9), 962-972.

Tsoy, D., Tirasawasdichai, T., & Kurpayanidi, K. I. (2021). Role of social media in shaping public risk perception during COVID-19 pandemic: A theoretical review. *International Journal of Management Science and Business Administration*, 7(2), 35-41.

Wang, Z., Oates, T., and Li, J. (2019). Deep Learning for Time Series Forecasting: A Survey. *IEEE Transactions on Neural Networks and Learning Systems*, 30(1), 201–214.

Wang, Z., Oates, T., and Li, J. (2019). Time Series Forecasting with Deep Learning: A Survey. *IEEE Transactions on Knowledge and Data Engineering*, 31(10), 1934–1947.

Wiranda, L., & Sadikin, M. (2019). Penerapan Long Short Term Memory pada Data Time Series untuk Memprediksi Penjualan Produk PT. Metiska Farma. *Jurnal Nasional Pendidikan Teknik Informatika*, 8(3), 184–196.

Witschge, T., et al. (2010). *Online Journalism: A Handbook for Digital News Production*. Routledge.

Woo, H., Cho, Y., Shim, E., Lee, J.-K., Lee, C.-G., Kim, S. H. (2016). Estimating influenza outbreaks using both search engine query data and social media data in South Korea.
*Journal of Medical Internet Research*, 18(7), e177.

World Health Organization (WHO). (2020). Coronavirus disease (COVID-19) pandemic. Retrieved from `https://www.who.int/emergencies/diseases/novel-coronavirus-2019`.

Qamar, R., & Zardari, B. A. (2023). Artificial neural networks: An overview. *Mesopotamian Journal of Computer Science*, 2023, 124-133.

Zhang, G. P. (2003). *Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model*. Neurocomputing, 50, 159–175.

Zhang, G. P. (2019). *Neural Networks for Time Series Forecasting*. In *Handbook of Neural Computing and Applications* (pp. 347–362). Springer, Cham.

Zahara, A., Sugianto, A. (2021). Multi-Layer Perceptron for Pattern Recognition: A Comprehensive Review. *Journal of Computer Science and Technology*, 36(2), 345-360. `https://doi.org/10.1007/s11390-021-0148-3`

Zhou, P., Yang, X. L., Wang, X. G., Hu, B., Zhang, L., Zhang, W., ... Shi, Z. L. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 579(7798), 270-273.