CHAPTER I
INTRODUCTION

1.1 Background

Air is an essential element for the survival of humans, animals, and plants. According to Dinas Lingkungan Hidup Kota Semarang, air is a mixture or collection of gases spread across the Earth's surface. The composition of good-quality air includes 78.08% nitrogen, 20.95% oxygen, 0.934% argon, 0.0314% carbon dioxide, 0.00182% neon, 0.000524% helium, 0.0002% methane, and 0.000114% krypton (Dinas Lingkungan Hidup Kota Semarang, 2020). Thus, air plays a crucial role in sustaining all living beings. On the other hand, air is not just a mixture of gases but an essential element that supports various biological and physical processes on Earth. From supplying oxygen for respiration to regulating the climate through heat distribution and cloud formation, the benefits of air are diverse and vital for sustaining life on Earth. Therefore, maintaining air quality is a crucial step to ensure that all natural processes supporting life continue to function properly (Dinas Lingkungan dan Kebersihan Kabupaten Mamuju, 2023).

Nowadays, air quality has deteriorated due to increasing industrial, transportation, and urbanization activities. Gases, particles, and compounds released into the atmosphere from human activities such as inefficient fuel combustion, agricultural or livestock waste burning contribute to declining air quality, known as air pollution. Air pollution is caused by two main categories: primary pollutants and secondary pollutants. Primary pollutants are directly emitted from sources, while secondary pollutants result from chemical reactions in the atmosphere. Over time, air pollution can lead to respiratory diseases, heart conditions, and even cancer (Dinas Lingkungan dan Kebersihan Kabupaten Mamuju, 2023; Dinas Lingkungan Hidup Kota Semarang, 2020).

Based on its components, air pollution consists of gaseous elements such as carbon monoxide (CO), sulfur dioxide (SO_2), nitrogen dioxide (NO_2), ozone (O_3), and ammonia (NH_3). Another component of air pollution is particulate matter, such as $PM_{2.5}$ and PM_{10}. $PM_{2.5}$ can penetrate deep into the lungs due to its aerodynamic diameter of less than 2.5 micrometers, allowing it to enter the bloodstream and increase the risk of cardiovascular diseases and respiratory disorders. Additionally, an increase in fine particle concentrations contributes to the formation of "haze," particularly in urban areas. Other air pollution components include volatile organic compounds (VOCs) and heavy metals, which pose long-term health risks when inhaled (Anandari et al., 2024; Ratnasari & Sofia Asharhani, 2021; Sakina et al., 2022)

To ensure that air quality remains healthy, the government has implemented various regulations and ambient air quality standards. In Indonesia, air quality regulations are stipulated in Peraturan Pemerintah Republik Indonesia Nomor 22 Tahun 2021 on Environmental Protection and Management and Peraturan Menteri Lingkungan Hidup dan Kehutanan Republik Indonesia on Air Pollution Standard Index (ISPU). These standards set the maximum concentration limits for various air pollutants, including $PM_{2.5}$, PM_{10}, SO_2, NO_2, CO_2, and O_3. For instance, the ambient air quality standard for $PM_{2.5}$ is set at 15 µg/m³ for annual averages and 55 µg/m³ for daily averages, which aligns with the stricter WHO guidelines. The enforcement of these regulations aims to control air pollution and protect public health, particularly vulnerable groups such as children, the elderly, and individuals with chronic respiratory diseases. With these standards in place, air quality monitoring is becoming increasingly important to ensure compliance with regulations and to serve as a foundation for designing more effective air pollution mitigation policies (Astrini et al., 2022; Firdaus et al., 2024).

Although regulations have been established, pollutant levels can still exceed the maximum threshold, as seen in DKI Jakarta. According to (IQair, 2025) Jakarta ranked 26th globally in terms of poor air quality, classified as "unhealthy." In 2017, Indonesia ranked 4th among the most polluted countries worldwide (IQair, 2025). The decline in ranking indicates changes in population activities, industrial operations, and private transportation usage. Nevertheless, continuous air quality monitoring is essential to identify pollution patterns through data analysis and to facilitate air pollution control efforts. One approach to air quality control is through prediction models. These models are designed to provide accurate air quality estimates for the future.

Artificial intelligence-based prediction models, particularly machine learning, offer a potential solution for forecasting air quality using historical data provided by the DKI Jakarta government’s air quality monitoring stations (SPKU) (Astrini et al., 2022; Kurniawan, 2017; Sakina et al., 2022; Sang et al., 2021; Wihayati & Wibowo, 2021). Machine learning excels at handling big data and extracting complex patterns from historical data, enabling it to identify non-linear relationships that traditional statistical methods struggle to capture. One of the most promising techniques for time series data analysis is Long Short-Term Memory (LSTM), a variant of Recurrent Neural Networks (RNNs) designed to handle long-term dependencies in sequential data. Using an LSTM model, air quality predictions can be made by considering historical data trends over specific time frames. In air quality forecasting, LSTM can recognize long-term trends and seasonal patterns that influence pollutant concentrations in the atmosphere (Mahendra et al., 2024).

In addition to LSTM, the Gated Recurrent Unit (GRU) model is also commonly used in time series modelling due to its simpler architecture while still being effective in handling sequential data (Furizal et al., 2024). GRU employs a gating mechanism that allows information retention over extended periods. GRU shares similar advantages with LSTM in capturing temporal patterns but features a more simplified architecture by merging the input and forget gates into one. This makes GRU computationally and memory-efficient during both training and inference compared to LSTM. Thus, both LSTM and GRU offer distinct advantages for time series modelling in air quality prediction in DKI Jakarta (Furizal et al., 2024; Handhayani, 2023; Yudiskara et al., 2023)

This study aims to predict air quality in DKI Jakarta using two deep learning models, LSTM and GRU, based on data from SPKU stations across DKI Jakarta. The dataset is obtained from the https://rendahemisi.jakarta.go.id website, covering a one-year period in 2024. Specifically, the selected monitoring stations include:

1. SPKU DKI1 Bundaran HI (Central Jakarta): A commercial area with high traffic and productivity;
2. SPKU DKI2 Kelapa Gading (North Jakarta): Represents a green area with relatively cleaner air;
3. SPKU DKI3 Jagakarsa (South Jakarta): Represents a green area with relatively cleaner air;

4. SPKU DKI4 Lubang Buaya (East Jakarta): represent about aviation activities and road traffic that can influenced air quality;
5. SPKU DKI5 Kebun Jeruk (West Jakarta): Close to industrial zones and main roads.

The data selection is based on its availability on the platform, an open-source website managed by Dinas Lingkungan Hidup, Pemprov DKI Jakarta, with support from Vital Strategies and Bloomberg Philanthropies. With high credibility, the dataset from this website serves as a foundation for ensuring the accuracy of air quality predictions.

This study is expected to develop a machine learning model capable of hourly pollutant concentration predictions, assisting policymakers in designing more effective mitigation strategies. Furthermore, this research aims to contribute to the development of modern, data-driven air quality monitoring technology.

1.2 Research Problem Identification

Based on the background described above, the identified problems are as follows:

1. The air quality in the DKI Jakarta has deteriorated due to industrial activities, transportation, and urbanization. Air pollution caused by gases, particles, and hazardous chemical compounds has a negative impact on public health and the environment in the region.
2. The Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models have great potential in time series data analysis. However, the effectiveness of these models in predicting air quality in the DKI Jakarta still needs to be tested and compared.

1.3 Problem Formulation

Based on the problems mentioned above, the research questions that can be formulated are as follows:

1. How can the Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) machine learning models be implemented to predict air quality in the DKI Jakarta?
2. Which model provides better performance to predicting air quality in the DKI Jakarta?

1.4 Problem Limitations

Based on the identified research questions, the author has defined the scope of the study to focus on addressing the identified issues. This research focuses on the implementation and comparison of LSTM and GRU models in predicting air quality in DKI Jakarta. The models will be trained using the expanding window approach in Time Series Cross-Validation. The data used consists of $PM_{2.5}$, PM_{10}, SO_2, NO_2, CO, and O_3, as these pollutants have a dominant influence on air quality. The data covers the period from January 2024 to December 2024, with predictions made for several hours ahead in January 2025. The data is sourced from five monitoring stations around the DKI Jakarta area.

Model evaluation will be conducted using metrics RMSE, MSE, MAPE, MBE and R^2. The prediction results will be converted into the Indeks Standar Pencemaran Udara (ISPU) to compare the prediction value with actual value (particles values). The best model selected based on its evaluation metrics value and prediction air pollutants result. Testing process used data from January 2025.

1.5 Research Objective

Based on the research problems presented, the objectives of this study are as follows:

1. To understand the performance of each model in predicting air quality in the DKI Jakarta.
2. To select the best model for air quality prediction in the DKI Jakarta.

1.6 Research Benefits

This research offers both theoretical and practical benefits. Theoretically, it contributes to a deeper understanding of deep learning models in time-series prediction case studies. Practically, it provides a direct impact on the research components. The theoretical and practical benefits of this study are as follows:

1. Theoretical Benefits:
   a. This research contributes to the advancement of science and technology in time-series-based prediction.
   b. It deepens the understanding of deep learning model architectures and their application in air quality monitoring case studies.
2. Practical Benefits:
   a. For the researcher, this study serves as a platform to explore ideas in the scientific field, enhance programming skills in developing accurate prediction models, and improve analytical abilities throughout the research process.
   b. For the public, this research provides air quality predictions for the next few hours, enabling individuals to make informed decisions about outdoor activities.

c. For the government, this study aids in decision-making and policy formulation when air pollution levels reach hazardous thresholds for public health.