CHAPTER I
INTRODUCTION

1.1 Background

Climate change has emerged as a global challenge affecting weather patterns across numerous regions of the world, characterized by a significant rise in the average surface temperature of the Earth. The latest report from the Intergovernmental Panel on Climate Change (IPCC) confirms that the most recent decade represents the warmest period ever recorded in history, giving rise to extreme weather anomalies [1]. This global warming phenomenon necessitates more precise air temperature monitoring and prediction in order to mitigate the resulting environmental impacts.

Surabaya, as the second largest metropolitan city in Indonesia, has experienced the direct consequences of this global warming trend, further exacerbated by the Urban Heat Island (UHI) effect [2]. The UHI phenomenon occurs when urban areas absorb and retain significantly greater amounts of heat compared to surrounding rural areas, owing to the predominance of construction materials such as concrete and asphalt that have replaced natural vegetation. The rapid expansion of built-up areas in Surabaya has substantially reduced green open spaces, causing heat to become trapped at the urban surface and raising local temperatures significantly relative to surrounding regions. Based on monitoring data from the Indonesian Meteorology, Climatology, and Geophysics Agency (BMKG), the maximum temperature recorded in Surabaya in 2023 reached 35.5°C during the months of October and November, with the perceived heat sensation considerably higher due to elevated atmospheric humidity [3].

These conditions carry serious implications for the quality of urban life and infrastructure management. Unpredictable temperature fluctuations directly impact the public health sector, including increased risks of dehydration and heat stroke, and influence electricity consumption patterns, particularly the sharp rise in air conditioning usage during periods of extreme heat [4].

Official meteorological station networks are typically sparse and predominantly situated in rural or open areas in accordance with World Meteorological Organization standards, and therefore fail to adequately represent urban microclimate conditions [5]. The forecasts produced by such platforms are generic in nature and are not specifically calibrated to account for the distinct characteristics of individual cities, such as the UHI effect particular to Surabaya. Furthermore, these platforms do not provide open access to their predictive models, rendering them incapable of being adapted for specific local analytical needs. The present study addresses this gap by developing a temperature prediction model based on local historical data from Surabaya, optimized and deployed through an interactive web application.

In conducting weather prediction, BMKG generally employs the Numerical Weather Prediction (NWP) approach, a physics-based atmospheric modeling method that simulates weather dynamics using a range of meteorological variables including air pressure, humidity, wind speed, and solar radiation [6]. This model utilizes observational data collected from various instruments such as weather radar, Automatic Weather Stations (AWS), and satellite data, which are subsequently processed using high-capacity computational systems [7].

Time series analysis constitutes a critical approach for modeling historical meteorological data patterns with the aim of projecting future conditions. Air temperature prediction enables relevant stakeholders to anticipate periods of extreme temperature. By leveraging daily historical data that embody seasonal patterns and long-term trends, time series methods are capable of extracting vital information concealed within apparently random weather data fluctuations [8]. Conventional statistical methods such as ARIMA (AutoRegressive Integrated Moving Average) are often limited in their capacity to handle complex weather data, particularly due to strict linearity assumptions and difficulties in capturing multiple seasonality patterns commonly encountered in climate datasets [9].

As a more robust alternative, the Prophet algorithm developed by Meta (Facebook) offers a more adaptive approach to meteorological data prediction. Prophet is specifically designed to handle time series data characterized by strong seasonal patterns, and is resistant to missing values and outliers, features that are commonly observed in weather sensor data [10]. A comparative study by Kwarteng et al. (2025) [11] demonstrated that Prophet is capable of outperforming ARIMA in terms of predictive accuracy, as measured by Mean Absolute Error, on datasets exhibiting dynamic seasonal fluctuations, owing to its flexible additive modeling structure.

Nevertheless, the superior performance of Prophet is not universally applicable across all data characteristics. The city of Surabaya, with its prevailing UHI phenomenon, presents specific temperature data characteristics: non-linear trend patterns resulting from continuous urban growth, high daily variance, and inconsistent seasonal pattern shifts between the dry and wet seasons. These characteristics may render Prophet's default parameters suboptimal. This is corroborated by empirical evidence presented by Km Priyanka Saini et al. [12], who reported that the default Prophet configuration frequently yields suboptimal results on volatile or highly seasonal data, and that its performance is highly sensitive to hyperparameter configuration. There is therefore a strong basis for arguing that the default Prophet model requires further optimization in order to perform effectively on daily temperature data from Surabaya, which is characterized by Urban Heat Island dynamics [13].

To substantiate this hypothesis, a preliminary experiment was conducted by constructing a Prophet model using default parameters on daily temperature data from Surabaya covering the period 2020–2024. The results of this initial experiment revealed that the default Prophet model yielded an RMSE of 0.8684°C, an MAE of 0.6602°C, and a MAPE of 2.33%, indicating that considerable room for improvement remains, particularly in capturing local trend variations. Prophet, which is theoretically robust and superior, had not yet reached its optimal performance when directly applied to the temperature data characteristics of Surabaya under Urban Heat Island dynamics. These suboptimal results provide a compelling basis for pursuing further optimization.

To address this gap, systematic hyperparameter optimization is required in order to identify the best-performing parameter configuration suited to the characteristics of Surabaya's temperature data. The Grid Search method represents an appropriate solution, as it performs an exhaustive empirical search across the entire hyperparameter combination space. By evaluating various combinations of changepoint_prior_scale, seasonality_prior_scale, and seasonality_mode values, Grid Search is expected to identify a configuration that enables the Prophet model to perform optimally on data exhibiting UHI characteristics [14].

Based on the foregoing discussion, this study aims to optimize the performance of the Prophet model through Grid Search hyperparameter tuning for air temperature prediction in Surabaya. By directly comparing the performance of the default Prophet model against the Grid Search-optimized model, this study seeks to demonstrate that the tuning process yields a meaningful improvement in predictive accuracy for temperature data with distinct local characteristics. The outcomes of this research are intended not only to produce a high-accuracy predictive model, but also to offer a practical contribution in the form of more precise temperature pattern insights for local government authorities and the general public in addressing the challenges posed by local climate uncertainty.

1.2 Problem Formulation

Based on the background described above, the problem formulation of this study is as follows:

1. What are the characteristics and patterns of daily average temperature data in Surabaya for the period 2020–2025, in terms of trend, annual seasonality, and weekly seasonality components?
2. How can an air temperature prediction model be developed using the Prophet algorithm?
3. What is the effect of hyperparameter optimization using the Grid Search method on the performance of the Prophet model?
4. How does the performance of the default Prophet model compare to that of the Grid Search-optimized model based on the evaluation metrics of RMSE, MAE, and MAPE?

1.3 Research Objectives

Based on the problem formulation above, the objectives of this study are as follows:

1. Analyze the characteristics and patterns of daily average temperature data for the city of Surabaya for the period 2020–2025, including trends and seasonality.
2. Build an air temperature prediction model using the Prophet algorithm with default parameters as the baseline model.
3. Optimize the Prophet model's hyperparameters using the Grid Search method to improve prediction accuracy.
4. Evaluate and compare the performance of the default Prophet model and the optimized model based on the metrics RMSE (Root Mean Square Error), MAE (Mean Absolute Error), and MAPE (Mean Absolute Percentage Error).

1.4 Research Benefits

This study is expected to provide benefits both theoretically and practically, as follows:

1. Contribute to the advancement of knowledge in the fields of prediction and time series analysis, particularly regarding the application of the Prophet algorithm to meteorological data.
2. Provide additional references and literature on the implementation of hyperparameter optimization techniques using Grid Search in the Prophet model.
3. Enrich empirical studies on air temperature prediction in tropical regions, particularly in Indonesia.
4. Provide a deeper understanding of the influence of Prophet hyperparameters on prediction accuracy for temperature data.

1.5 Research Limitations

The following limitations are applied in this study:

1. The data utilized in this study consist of daily average temperature records (Tavg) obtained from timeanddate.com, based on observations from the Surabaya/Perak Meteorological Station, covering the period from January 1, 2020 to December 31, 2025.
2. The prediction model employed is Prophet, developed by Meta (Facebook).
3. The hyperparameter optimization method applied is Grid Search, with a focus on three primary parameters: changepoint_prior_scale, seasonality_prior_scale, and seasonality_mode.
4. The values of changepoint_prior_scale tested are [0.05, 0.1, 0.3], the values of seasonality_prior_scale tested are [5, 10, 20], and the seasonality_mode values tested are ["additive", "multiplicative"].
5. Model evaluation is conducted using the RMSE, MAE, and MAPE metrics, with the dataset partitioned into training and testing subsets, wherein the testing set comprises the final 365 days of the dataset.
6. This study does not incorporate exogenous variables such as humidity, rainfall, wind speed, or air pressure into the prediction model.
7. The interactive web application developed in this study is operated locally using the Streamlit framework and has not been deployed to a public server; its accessibility is therefore currently limited to the local development environment.