CHAPTER I
THE PROBLEM

1.1 Background

Congestion is a real problem faced by many countries in the world, especially developing countries [2]. This is due to the need for humans to mobilize for their daily activities. Which is where the decision to buy a private vehicle is also supported by an increase in one's income. However, this increase in private vehicle ownership has also resulted in traffic congestion in many places. Traffic congestion also causes various effects such as air pollution, more fuel usage, and longer travel time [3]. Long travel durations due to congestion will certainly have a greater impact on air pollution. Which, air pollution can also impact vegetation and disturb natural breathing processes [4]. Since the impact of traffic congestion if not handled seriously can cause problems in health aspects such as physical and mental health of the community [4–6], the traffic congestion will also lead to material costs. In Chittagong city, Bangladesh, the losses caused by long travel times were 2.01 million dollars per day [7]. Even China, a mature country, still received a loss of 31.5 billion yuan in 2019 for the Beijing-Tianjin-Hebei region [8]. One way to deal with congestion is to educate people to switch to public transportation, such as buses. Of course, education about using public transportation such as buses must be followed by good bus service facilities from bus service providers. There are various ways to increase user satisfaction when using the bus including providing information about existing bus routes, providing information about the location of bus stops, providing clear information about the position of each bus, to provide bus travel time prediction.

Predicting bus travel time can improve bus user satisfaction [9]. This is able to happen because every user can know the exact time of their travel. Prediction of bus travel time will be easier in areas that have dedicated bus lanes [10]. This is because bus-only lanes make the bus movement environment more static so it is easier to predict [11]. Predicting bus travel time will be more difficult if it is done in areas that do not have bus lines. This is because the bus will run in the same lane as private vehicles. So this condition causes the bus travel environment to be more dynamic. Predicting bus travel time will be more difficult if it is done in areas that do not have bus lines [11].This is obviously due to the dynamic conditions. So

there needs to be a special approach to predict bus arrival times in dynamic conditions. In predicting travel time, there are several methods that are common and have been used by the community. The method referred to is using google maps in predicting travel time. Several methods are common and have been used by the community to predict travel time. One method is using Google Maps to predict travel time and congestion. In predicting travel time and congestion, Google Maps uses the user's Global Positioning System (GPS) data as the source of the dataset. This GPS data is taken in real time by Google Maps from the user's smartphone. The use of GPS data as a real-time data source will give a more definite user position. So that the travel time prediction process could provide almost accurate or even accurate prediction results. However, Google Maps predictions cannot follow the bus routes that have been set in each place. This is due to the fact that in making predictions, Google Maps will prioritize the shortest and fastest route that can be taken. In other words, Google Maps has limitations in terms of selecting the route you want to predict. To answer this problem, there was a previous study that predicted bus travel time using machine learning algorithms with GPS data as the dataset. To answer this problem, there is previous research that was conducted by Ankit Taparia and Mike Brady [1]. In this research, Ankit and Mike predicted bus travel time using machine learning algorithms with GPS data as the dataset. Machine learning algorithms utilized include Linear Regression, Artificial Neural Networks (ANNs), and Long Short Term Memory (LSTMs) with datasets collected at Dublin City Council from November 6, 2012 to November 30, 2012 [1]. The algorithms used are also carefully selected. Linear Regression was chosen because it is a simple algorithm and is helpful in predicting supervised quantitative data [12]. Artificial Neural Networks (ANN) are utilized because to their robust methodologies for gathering and modeling non-linear data with intricate interactions between inputs and outputs [13]. While recurrent neural networks (RNNs) have short-term memory issues, Long Short Term Memory (LSTM) networks are a type of RNN that can handle lengthy input sequences and forecast time series [14]. However, the limitation of this research is that prediction can only work well when applied to Dublin City. This is because the dataset utilized is only a dataset from Dublin City.

Predicting bus travel time in a dynamic environment has never been impossible. With the prediction method that has been made in previous studies, predicting bus travel time in the city of Bandung can be conducted. In this research, the method of predicting bus travel time with XGBoost, ANN, and LSTM algorithms will be implemented on a dataset that will be generated according to bus routes in Bandung City. In building the dataset, the route selected is the Baleendah - Bandung Elec-

tronic Center route. Predictions will be made for 3 bus stops that have been selected based on the frequency of passenger exchanges. Which after the prediction is performed, the prediction results will be verified with the actual situation. So that the Root Mean Square Error (RMSE), Mean Absolute Error (MAE), R-squared (R^2), Mean Absolute Percentage Error (MAPE), and Median Absolute Error (MdAE) values can be seen. With this, the performance of the prediction model can be seen based on the margin of error. Which is, the smaller the difference, the better the prediction. This research not only focuses on the small error value of a prediction but also focuses on creating a new dataset that matches the actual city of Bandung data.

1.2 Theoretical Framework

The technique used to make predictions is a prediction technique using computational media. In this case, computing will be run through cloud-based computing services so that the computing system will be carried out on the service provider's server. Computing activities are carried out using computing services so they do not burden the hardware devices used. Which this step can also provide flexibility in performing the program. In programming the model, the algorithm that will be used is the machine learning algorithm as the main comparison algorithm. Machine learning algorithms are often used as algorithms for building prediction models because they have a better ability to generalize. It also has a more specific network architecture.

In this research, the algorithms that will be used are XGBoost, ANN, and LSTM. These algorithms are utilized because the ANN can model non-linear data with complex interactions between inputs and outputs. Meanwhile, LSTM is utilized because it is an interpretation of RNN with the ability to process long series of inputs and predict time series. In addition, there are also XGBoost algorithms that are also utilized as comparison materials. This is because XGBoost has a simpler concept. Therefore, XGBoost will act as a feasibility test of the complexity of deep learning itself to be applied.

1.3 Conceptual Framework

This research utilizes a prediction system that incorporates three algorithmic types: Extreme Gradient Boosting (XGBoost), Artificial Neural Network (ANN), and Long Short-Term Memory (LSTM). The model training approach was car-

ried out with a manually gathered dataset based on the Baleendah-Bandung Electronic Center bus route. This dataset contains comprehensive information, including weather, natural disaster, event, and arrival time of the bus at three designated stops: Baleendah Bus Stop, Alun-alun Bus Stop, and Bandung Electronic Center Bus Stop. The algorithm is anticipated to learn bus travel time using this dataset in order to generate precise predictions.

The bus travel dataset on the Baleendah-Bandung Electronic Center route is used for the model testing procedure, but the data collection time period differs from that of the dataset used for training. The model's success in making predictions will be measured by the error rate obtained during testing. The discrepancy in errors is quantified using evaluation metrics like RMSE, MAE, R^2, MAPE, and MdAE. These five criteria are used to provide a more thorough assessment of the model's correctness and its capacity to manage novel inputs.

1.4 State of The Problem

According to the above background, there are several problems that will be solved. These are some of the problems that will be solved in this research:

1. Predicting the travel time of buses at Bandung city which has a dynamic traffic environment.

2. Determine the suitable algorithm for bus travel time prediction.

1.5 Objective

Based on the above background, several objectives can be formulated. The following are some of the objectives that have been formulated for this study:

1. Creating a dataset that reflects the actual conditions in the Bandung city.

2. Find the algorithm that gives the smallest error for the dataset that has been created.

1.6 Hypothesis

This research hypothesizes that predicting bus travel time using a dataset based on the conditions of the Baleendah-Bandung Electronic Center bus route in Bandung city will yield relatively low RMSE, MAE, R^2, MAPE, and MdAE values.

This is because the dataset used reflects the actual conditions of Bandung city. Consequently, the prediction of bus travel time will not experience delays, and if delays occur, they will not exceed 5 minutes.

The connection between the data used and the prediction model applied was also taken into consideration in this research. As the primary input that influences the learning process of the model, the data contained in the dataset acts as the main input. On the other hand, the prediction algorithm seeks to provide predictions of the amount of time it takes for a bus to travel based on patterns discovered in the data. The way data and algorithms work together to produce a minimal error number is demonstrated by this connection.

1.7 Assumption

The dataset used in this study consists of bus travel data collected directly from the Baleendah-Bandung Electronic Center bus route, which is thus assumed to represent the dynamic conditions of this route accurately. The data preprocessing performed in the learning program effectively cleanses the data of any empty values. This data cleansing is expected to yield a dataset capable of producing a prediction model with low RMSE, MAE, R^2, MAPE, and MdAE values.

1.8 Scope of Work

To limit the space for discussion, the scopes of work in this research can be formulated as below:

1. Focuses solely on predicting bus travel time, disregarding seat availability on a bus.
2. Compares predictions across selected algorithms only, not utilizing all available algorithms.
3. Not all bus stops are included in the predictions; only four stops are covered: Baleendah, Alun-alun (outbound), Bandung Electronic Center, and Alun-alun (return).
4. Not taking into account delays that occur in the prediction process.
5. Only develop a prediction algorithm without creating an application to predict bus travel times.

1.9 Importance of the Study

This study aims to estimate travel time, the predictions resultse evaluated using RMSE, MAE, R^2, MAPE, and MdAE values. Bus travel time prediction could be achieved by utilizing machine learning as a medium. The predictions made by machine learning are designed with consideration for the time the bus is in transit, weather conditions, natural disasters, and current events.