CHAPTER I

INTRODUCTION

1.1 Background

The development of wireless technology, particularly for the Internet of Things (IoT) ecosystem with remotely connected devices, has been unprecedented over the past few years. By 2027, it is estimated that 30.2 billion IoT devices (including short-range IoT, cellular, and other long-range IoT segments) will be active worldwide [1]. One of the key aspects in facilitating efficient and reliable connectivity for IoT is advanced radio modulation techniques, which include the spreading factor [2], [3]. The spreading factor plays a crucial role in determining spectrum efficiency, battery life, and communication range for various types of IoT devices, ranging from small sensors to sophisticated mobile devices [4], [5]. The low power wide area networks (LPWAN) paradigm is gaining a lot of momentum in the field of massive Internet of Things (IoT) for its peculiarity of providing wide-area coverage while having low power requirements for transmission [6]. In this context, the use of machine learning-based classification techniques is increasingly important to dynamically optimize the setting of spreading factors according to changing networks and environmental conditions. Recent research and development highlight how machine learning algorithms can predict and adjust spreading factors with high precision, enhancing network efficiency, and providing better connectivity experiences for all IoT devices in this increasingly connected digital era [7], [8].

Research conducted by Christos John Bouras et al. [9] investigates the use of Low Power Wide Area Network (LPWAN) technology, specifically Long Range (LoRa) communication, in their study. Their research focuses on improving the determination of the spreading factor (SF) in LoRa networks using machine learning (ML) models to enhance energy consumption and data delivery ratios. The authors examine three methods of SF assignment: Random SF Assignment, Adaptive Data Rate (ADR), and ML-based SF Selection. They develop a simulation environment using OMNeT++ and ML libraries, training and evaluating various ML models such as k-Nearest Neighbors (k-NN), Naïve Bayes, Support Vector Machines (SVM), and Decision Trees. The dataset is balanced using SMOTE-NC.

The results show that ML-based methods generally outperform random SF assignment but are slightly less effective than ADR. The k-NN model achieves the highest performance metrics. Although ADR's dynamic power adjustment provides better performance in some cases, ML methods are promising but require optimization to fully match ADR. Integrating ML into LoRa networks is challenging, thus presenting a promising area for future research.

Toni Perkovic et al. [10] conducted research on the application of machine learning (ML) techniques to improve indoor localization in LoRaWAN networks. Their research focuses on utilizing signal strength variations, particularly RSSI (Received Signal Strength Indicator) and SNR (Signal-to-Noise Ratio), to determine the position of devices in indoor environments. The methodology involves collecting data from several LoRaWAN gateways, with RSSI and SNR values as the primary inputs to the ML model. The researchers chose neural network (NN) models for the localization task and experimented with various hyperparameters to optimize the model configuration. The results show that the best-configured model achieves an accuracy of 98.8% on the test set with a learning rate of 0.01 over 100 epochs. SF7 is identified as the optimal spreading factor for high localization accuracy, highlighting the importance of appropriate SF placement to reduce data overlap and improve precision. Evaluation metrics, including confusion matrices and accuracy plots, confirm the model's effectiveness in distinguishing indoor locations based on signal strength variations. However, the research has limitations, such as requiring too many gateways and needing to be applied in more realistic areas with more varied distances. In [11] also researched LoRa localization using machine learning; however, the research was conducted only in an outdoor area (a sports oval with an area coverage of 30,000 square meters) using RSSI, SNR, and SF as features, and achieved 50% better accuracy with a fingerprint-based direct location estimation approach.

Christos Bouras et al. [12] conducted research on the application of supervised machine learning (ML) techniques to optimize the allocation of spreading factors (SF) in LoRa networks. The goal is to enhance energy efficiency and connectivity by using ML algorithms for SF assignment, thereby improving the performance and reliability of IoT communications. Simulations were conducted using the FLoRa simulator, which accurately represents various aspects of actual LoRa networks. Data were collected by running simulations with various network configurations. The main features used to train the ML models were transmission power (TP), total energy consumed, and total packets sent. Three supervised learning algorithms were evaluated: k-Nearest Neighbors (k-NN), Naïve Bayes, and Support Vector Machines (SVM). The classification problem was set to predict the optimal SF (ranging from 7 to 12) for data transmission in a LoRa network. The k-NN model achieved high accuracy (95% with k=3) and performed well in urban and suburban scenarios. Naïve Bayes showed adequate performance with an accuracy of around 80% but was less effective than k-NN and SVM. SVM showed the highest accuracy, especially in urban scenarios (96.79%). The study concludes that supervised learning techniques, particularly k-NN and SVM, are effective in optimizing SF allocation in LoRa networks. These ML models significantly improve energy efficiency and connectivity, making them suitable for low-cost and low-power IoT applications. However, the research did not vary power input and has not yet attempted real-life application scenarios.

Several machine learning methodologies have been investigated and implemented to enhance LoRa performance [13]. However, there are some issues in the implementation of the research. As mentioned, experiments on real field data have not been conducted and are limited to simulation tools. The observation area in the research was limited to only 480 meters [14]. Furthermore, studies on LoRa performance have predominantly been carried out in sub-tropical areas [15], [16], [17]. Then in the study [12], no research has been conducted on classification to determine the power value and in the classification process the SF value only gets the highest accuracy of 96.79% without explaining the testing time and training time. So, building an accurate classification model and having a short testing time is very important in developing LoRa communication to obtain the most appropriate settings for the SF value and power value in LoRa communication, which affect network performance, coverage, data rate, and energy consumption.

This research provides a detailed comparison of several machine learning models, evaluating their performance in terms of prediction accuracy, computational complexity, and scalability. This comparative study utilizes a comprehensive dataset of LoRa network parameters and real-world scenarios to validate the effectiveness of these models. The dataset was collected from rice field areas in Banyumas, Central Java, Indonesia, which has a tropical climate. The area is open and near residential neighborhoods. The area contains trees and there are activities of living things that can be obstacles and can also affect the quality of signals or data sent by the transmitter to the receiver. The average data collection time is in the morning from 7-10 am and in the afternoon from 3-5 pm with the distance from the receiving antenna to the transmitting antenna varying from 100 meters to 800 meters with an increase of 50 meters at each point. By systematically analyzing the strengths and weaknesses of each machine learning model, this study aims to offer valuable insights for researchers and practitioners in the fields of IoT and LPWAN. The findings can guide the development of more efficient model of classification SF and power selection algorithms, ultimately enhancing the performance and reliability of LoRa networks.

1.2 Problem Statement

Based on the background presented, the research problem statements in this study are:

1) How is the proper classification process of the spreading factor and power in node-to-node LoRa communication?
2) How is the performance comparison of the machine learning algorithms used in spreading factor and power classification?

1.3 Research Objectives

Based on the problem statement above, the objectives of this research are as follows:

1) To understand the process of classifying the spreading factor and power in node-to-node LoRa communication.
2) To determine the machine learning algorithm's performance in carrying out spreading factor and power classification.

1.4 Scope Of Work

To narrow the scope of discussion in this research, the problems formulated are as follows:

1) The dataset used for this study is node-to-node LoRa data taken in a rice field area in Banyumas, Central Java, Indonesia.
2) Data collection time is in the morning around 7 - 9 am and in the afternoon around 3 - 5 pm.
3) SF values on dataset from 7-12.
4) Observation distances of 100 m - 800 m.
5) Power setting 0 dBm – 20 dBm.
6) The size of the packets sent varies, namely 100 bytes, 175 bytes, and 250 bytes.
7) The number of packets sent for each distance with variations in device setting parameters is 100 packets.
8) The LoRa module used is Mappi32.
9) The bandwidth used is only 125 kHz.
10) The default coding rate setting is at 5.
11) The classification process uses the KNN, Random Forest, and Decision Tree machine learning methods.
12) The parameters used to measure the performance of the classification system are accuracy, training time, and testing time.

1.5 Hypothesis

The use of machine learning to determine the spreading factor and power in LoRa communication can enhance network efficiency by achieving an optimal balance between signal range, data transfer speed, and energy consumption. By classifying the spreading factor values and power values, the performance of LoRa can be optimized as the spreading factor and power used will match the specific needs.

1.6 Research Timeline

The research timeline helps organize, set deadlines, and progress monitoring. Table 1.1 shows the research project, from the initial idea to the final report.

Table 1. 1 Research Timeline

Activity	2023							2024							
	Jun	July	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
Literature Review															
Collect Dataset															
Pre-processing Dataset															
Creating System															
Writing Section I-III															
Training and Testing System															
Evaluation System and Analysis															
Report Writing Section IV															

1.7 Structure of Thesis

This research is divided into five different parts which will be explained as follows.

- **CHAPTER 1: INTRODUCTION**
  This chapter discusses the background of the research, problems in the field, related research, problem statement, research objectives, and scope of the research.
- **CHAPTER 2: BASIC CONCEPTS**
  This chapter provides basic information for this thesis, including an explanation of LoRa and spreading factor, the LoRa module used, the dataset balancing process, and the machine learning method that will be applied in this research.
- **CHAPTER 3: SYSTEM DESIGN**
  This chapter explains the system model starting from the dataset collection process, research flow, and how the simulation works in the algorithm, including the parameters and variables used in the thesis.
- **CHAPTER 4: SIMULATION RESULT AND ANALYSIS**
  This chapter discusses the results of the final simulation task consisting of classification metrics, confusion matrices, and learning process time.
- **CHAPTER 5: CONCLUSION AND FUTURE WORKS**
  This chapter provides the conclusion of this thesis and the recommendations for future works.