CHAPTER 1
INTRODUCTION

1.1 Background

PT Kereta Api Indonesia, or PT KAI (Persero), as a train operator, is facing the challenge of fulfilling the increased public demand for rail transportation modes in the aftermath of the Covid-19 outbreak. It is indicated by the increase in the revenue of the company from passenger and freight transport services in 2021 until 2022 [1]. It is essential for PT KAI to guarantee that delivered services are safe, reliable, comfortable, and timely. The operation and maintenance of railway systems is one factor that affects the safety and availability of railway operations [2]. With increased railway operations, the usage lifetime of railway equipment will rise, thereby increasing the probability of equipment failure [3]. As a result, maintenance is required to keep an equipment's performance in accordance with its function. The maintenance methods consists of 3 types [4], namely preventive maintenance which is carried out periodically/periodically, reactive maintenance which is carried out to correct disturbances that occur in equipment and predictive maintenance (PdM) which aims to predict failures in equipment or operating machines and decide when maintenance is carried out. Nowadays, PT KAI (Persero) only implements preventive and corrective maintenance on all railway equipments.

One of critical equipment in the railway signaling system is railway point machine (RPM) that consist of electric motor and mechanical assembly with the function of moving, detecting, and locking the position end of the point tongue either individually or following the direction of the route formed [5]. The existing maintenance conditions of RPM equipment are 2-weekly preventive maintenance and annual overhaul maintenance [6]. 2-weekly preventive maintenance is carried out by doing visual inspections and some listed activities in maintenance checklist to confirm that the RPM can operate properly. While overhaul maintenance is performed, all components in the RPM equipment are disassembled, several component replacements are made, components that are not replaced are cleaned and then reassembled, after that quality check is performed following re-assembly using the RPM test-bench [6].

The biggest problem with existing RPM maintenance conditions is that there is no monitoring system on RPM equipment, thus maintenance is still done on a time basis, particularly every two weeks and overhaul annually. All RPM operating areas receive 2-weekly preventive maintenance, both in areas of busy train services, such as commuter operations, and areas of low train services, such as villages only passed by intercity operations [6]. Excessive maintenance on RPM equipment that is infrequently used will result in inefficient maintenance costs [7]. In addition to the problems with annual overhaul maintenance, the selection of which RPM is the priority target of maintenance is still restricted since overhaul implementation can only be done at UPT Balai Yasa Sintel and LAA Bandung, as well as the limited overhaul budget. The RPM overhaul maintenance is usually performed reactively, for example, if RPM equipment is dismantled as a result of the government replacing new RPM equipment, or when the RPM failures, or if the station operator or maintenance team reports damage to the RPM equipments [8].

A predictive maintenance (PdM) system is a maintenance method that PT KAI (Persero) has not implemented yet for maintenance solution for its assets. Based on BS EN 13306: 2017 [9], the implementation of predictive maintenance involves use of a sensor that monitors the equipment, and the data from the equipment condition monitoring is utilized as a decision-making resource in maintenance operations. PdM has the best maintenance performance from the optimization of costs and the number of maintenance frequencies [4] as shown in Figure 1.1.

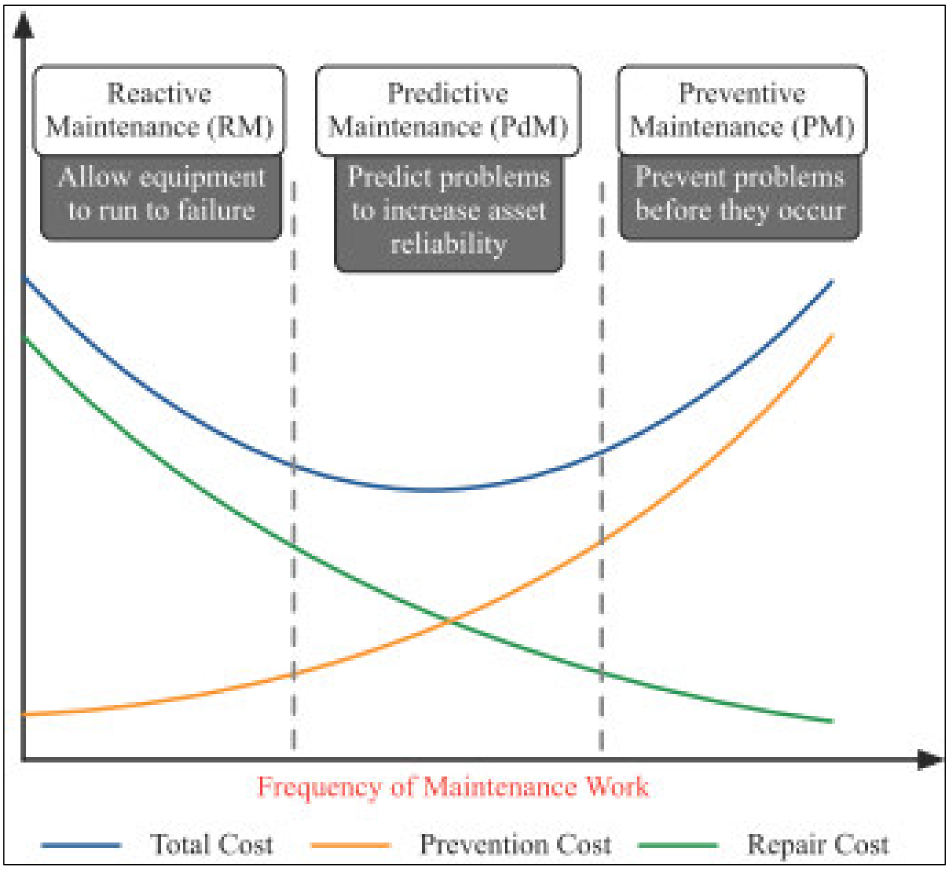

Figure 1.1 Frequency and cost of maintenance work curve [4]

Figure 1.1 explains that reactive and preventive maintenance has greater maintenance total cost than predictive maintenance because reactive maintenance has high repair cost, and the preventive maintenance has high preventive cost. In PdM, maintenance is performed before a service failure occurs and when the condition of the equipment actually requires maintenance. The equipment's condition is determined by analyzing the monitoring data. PdM is also the highest level of maintenance with the highest uptime and overall equipment effectiveness (OEE) value as illustrated in Figure 1.2, compared to other types of maintenance such as reactive maintenance, periodic maintenance, and proactive maintenance [10].

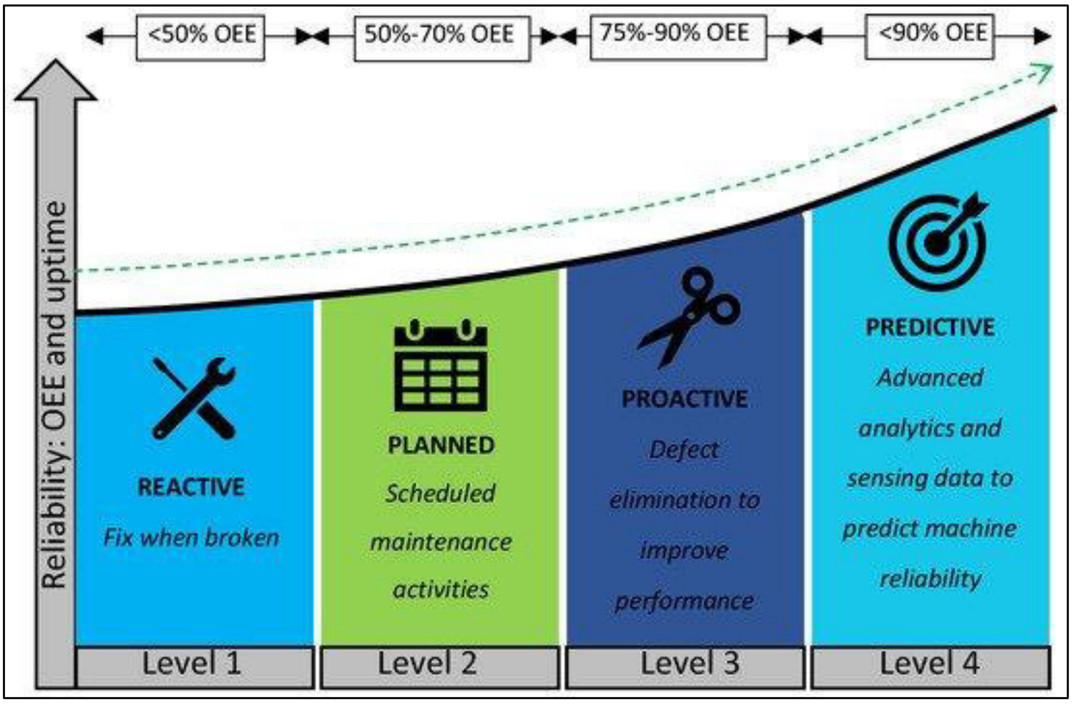

Figure 1.2 Comparison of OEE and uptime in maintenance types [10]

Figure 1.2 shows that PdM has more than 90% of OEE value and has the highest uptime which means highest productive time in operation compared to the others type of maintenance.

In predictive maintenance system, the maintenance decision comes from sensor data that installed for monitoring the equipment. There are several sensors that can be used for RPM equipment monitoring such as voltage [11], current [11], force [11], acoustic [12] and vibration [13]. Asada et al [11] found that current sensor is the most applicable in practical level because current sensor installation could not affect the operation and could be installed in power supply which is inside the equipment room in station area, not in the RPM equipment or outdoor area. Dhomad and Jaber [14] also found that current data of electric motor has high accuracy in representing of motor condition.

The collected current measurement from current sensor will produce different characteristic data for each RPM equipment condition. In predictive maintenance system, the collected current data needs to be analyzed for classifying the RPM equipment condition. There are two classification methods in data-driven based condition classification, which are statistical and machine learning methods [15]. Table 1.1 compares the advantages and limits of statistical and machine learning methods.

Table 1.1 Advantages and limitations of statistical and machine learning methods

Data-driven based Condition Classification	**Advantages**	**Limitations**
Statistical methods	- Easy to understand (noncomplex calculation) [16] - Lower computational cost [16]	- Relies on the assumption that parameters have a known distribution, which may approximate the true behaviour. - Does not consider the non-linearity of the data [17] - Time invariant, while most of the real processes are time-varying [18]
Machine learning methods	- Possible to detect new issues or faults with insufficient data [15] - Can describe very complex and non-linear systems with great accuracy in defect identification [18] - Capable of diagnosing, predicting failure, and calculating the lifetime of equipment [19]	- A large volume of training data is necessary for some ML methods [20] - High computational resources [21]

As listed in Table 1.1, the machine learning method has more benefits than the statistical method. In addition, in the era of Industry 4.0, artificial intelligence such as machine learning methods could help humans in decision making and make the maintenance job more efficient with high accuracy result [10].

There are some previous studies about predictive maintenance in railway point machine (RPM) equipment. T. Asada et al [11] developed the RPM equipment condition monitoring system which focused on three measurement parameters, namely the measurement of force, voltage, and electric current. They used the k-means method for clustering the results of the three measurements. Current and voltage data would be converted into electric power. The current and voltage parameters were measured using a Hall-effect current transducer and a voltage transducer. The force measurement had a better coefficient value than the electric power parameter for representing the RPM condition especially in overdriving condition, but the electric power measurement was more appropriate to be carried out at a practical level because it was safer and did not have the potential to disrupt the existing operation of the RPM equipment. The feature extraction method that was used, the Discrete Wavelet Transform (DWT) method. For fault detection and diagnosis, they used the pattern recognition Support Vector Machine (SVM) classification method. The study showed the comparison between linear and rbf kernel of SVM classification using power data of RPM equipment, which both methods got 1.0 cross validation accuracy. Wenjing Jin et al [22] made the RPM equipment condition monitoring system which was carried out by measuring voltage and current signals remotely, namely the on-board condition monitoring (OBCM) tool. The experiments were done in RPM test-bench. The current and voltage data would be transformed into power data which will be used as time-series data monitoring. The power data then would be extracted into some features which these features would be used as datasets for unsupervised classification methods. This study used two unsupervised methods, namely the Self-Organizing Map-Minimum Quantization Error (SOM-MQE) algorithm and the Principal Component Analysis T^2 (PCA-T^2) algorithm. SOM-MQE and PCA-T^2 methods could be used to detect performance degradation and evaluate equipment on several levels by putting an experience-based threshold to set the failure condition. The research focused on SOM-MQE method for classifying the RPM equipment because it had better characteristics for fault diagnosis which had accuracy result 0.81. Jonguk Lee et al [12] used ML applications for fault detection and diagnostics on RPM equipment from sound analysis obtained from audio sensors. The results of audio sensor readings were processed using mel-frequency cepstrum coefficients (MFCCs) for feature extraction and dimension reduction. These features would be used as input dataset for Classification-Support Vector Machine (C-SVM) algorithm using rbf kernel. The study found that the accuracy of C-SVM with 133 features (0.94 accuracy) is quite close to that of C-SVM with 720 features (0.97 accuracy). Jaewon Sa et al [23] used electric current data for classifying two condition of RPM equipment, namely good condition and bad condition that need to be replaced. The current time-series data was evaluated using subsequence and full sequence shapelet algorithms for time-series extraction and classification. The results of both methods were also compared to dynamic time warping (DTW) method. The study obtained that shapelet algorithms have much better accuracy (more than 0.9 accuracy) than DTW algorithm (0.6 accuracy) for classifying the binary RPM equipment condition.

The position of this study is to focus on developing a prototype of PdM system for RPM equipment in the area of PT KAI (Persero). The monitoring equipment uses an electric current sensor which is the most applicable in RPM monitoring implementation and has high accuracy for representing the RPM condition [11]. The current measurement results will produce different measurement data for each RPM equipment condition. The three conditions that will be monitored are normal condition, warning condition and failure condition. The collected time-series current data would be extracted into some features as dataset for machine learning modelling. This study will use three types of ML algorithms. They are Support Vector Machine (SVM), Random Forest (RF) and Artificial Neural Network (ANN) which are ML algorithms of the supervised learning type for classification purposes with the most use for PdM system applications [10]. The classification results of three ML models will be evaluated using the confusion matrix method [24] and compared to which ML model has the best accuracy in classifying the RPM equipment condition.

1.2 Research Objectives

The purpose of this research is to create a prototype of a predictive maintenance (PdM) system for the railway point machine (RPM) equipment with the following details.

1. Implementing an electric current sensor as a condition-based monitoring tool for the RPM equipment, with the results of sensor readings can represent the condition of RPM equipment.
2. Implementation of optimal classification machine learning (ML) algorithm in the for classifying the RPM equipment conditions.

1.3 Problem Formulation

The main requirement in a predictive maintenance (PdM) system for railway point machines (RPM) equipment is a simple system architecture, valid data acquisition process and high accuracy classification result.

The formulation of the problems in this study is as follows.

1. The electric current sensor used as a monitoring tool for RPM equipment can monitor RPM equipment operation and does not interfere with existing operations.
2. The selection of features that extracted from current-signal have relevant relationship to the classification of RPM equipment conditions.
3. The multi-classification machine learning (ML) algorithms that used in PdM system has high accuracy result with value more than 0.9.

1.4 Problem Limitation

The limitations of the problem in this thesis proposal include:

1. Related studies in this research are advanced embedded systems, intelligent systems and statistical learning and optimization.

2. The current sensor that is used in this research is the WCS1700 Hall Effect Base Linear Current Sensor that is connected to NodeMCU ESP8266 as embedded system.
3. Multi-classification machine learning (ML) algorithms that will be used for predictive maintenance (PdM) system are Support Vector Machine (SVM), Random Forest (RF) and Artificial Neural Network (ANN).
4. The railway point machine (RPM) type that will be used in this research is the NSE type of RMP that comes from *Nederlandse Machinefabriek Alkmaar* (NMA) manufacturer.
5. The working parameter to be analyzed is the dc electric current.
6. RPM equipment classification conditions are limited to:
   a. Normal condition when the RPM operates as its function in good condition.
   b. Warning condition when the friction assembly of RPM has lack of grease and it need to be maintained by cleaning and replacing the new grease.
   c. Failure condition when the RPM could not move into the end position.
7. The software that will be used in this research are Arduino IDE 1.8.19 software for programming the current sensor and Jupyter Notebook 6.4.8 software for programming the classification machine learning algorithms.
8. The research location is the UPT Balai Yasa Sintel dan LAA of PT KAI (Persero).

1.5 Hypothesis

The electric current of RPM equipment as the observation parameter could give good representation of RPM equipment condition. The current sensor reading will be the input for classification machine learning model for PdM system. The accuracy value of using SVM algorithm can reach higher than 0.8. Besides that, the accuracy value of using RF and ANN algorithms can reach higher than 0.9 [10]. Therefore, the hypothesis raised is that classification machine learning could identify the RPM equipment condition with high accuracy value more than 0.9.

The independent variable of this study is the electric current value of the RPM equipment. While the dependent variable of this study is the classification of RPM equipment conditions which are normal condition, warning condition and failure condition.

1.6 Research Methods

In this study, the focus of the research method is on the machine learning algorithm used, where the ML modelling will use a supervised machine learning type method [25]. Where the machine learning modelling uses training data that has been labelled with the three classes of RPM equipment conditions or called multiclassification supervised learning. This research will use the three most popular supervised learning algorithms used in equipment condition classification for predictive maintenance, namely SVM, RF and ANN [10][26]. The proposed method for the three algorithms is as follows:

1. Linear Support Vector Machine (SVM) Classifier

SVM is a popular ML method for classification and regression tasks due to its accuracy [26]. It excels at separating different classes of data precisely by determining the optimal point. In multiclassification problems, SVM creates n-dimension hyperplanes that divide data ideally into n groups/classes which has high accuracy in n-multiclassification for multi-variable datasets [12][26]. SVM classifier has several types of kernels that can be used such as linear, radial basis function (rbf), quadratic, cubic, polynomial, and fine gaussian [27]. Asada et al [12] found that linear and rbf kernel performed well in multiclassification for identifying the RPM equipment conditions using some features that came from electric power data. Both methods had the same accuracy value but compared to rbf kernel, linear kernel is much simpler and easier to implement. Linear SVM is used for linearly separable data, which means that if a data set can be classified into two or more classes using a straight line, it can be classified into two or more classes [28]. Figure 1.3 illustrates how linear SVM works in performing classification for multiclass.

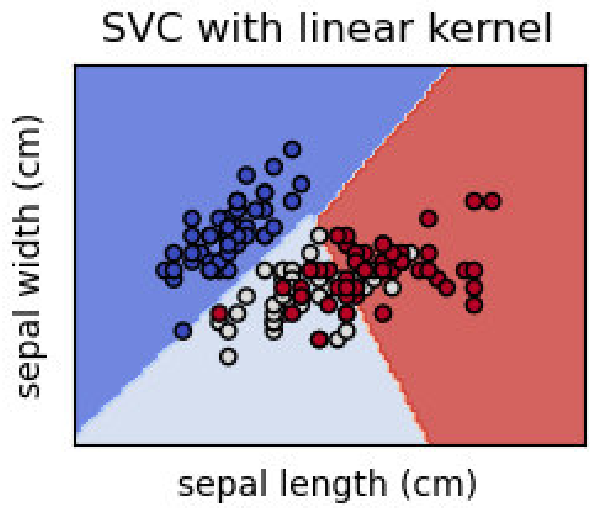

Figure 1.3 Support vector machine classifier with linear kernel [28]

As shown Figure 1.3, linear SVM do the classification by using linear decision boundaries between the variables of datasets. When fitting the SVM model with linear kernel, there is one parameter that need to be considered, which is the parameter C, common to all SVM kernels, trades off misclassification of training examples against simplicity of the decision surface. A low C makes the decision surface smooth, while a high C aims at classifying all training examples correctly. While Figure 1.4 illustrates how SVM with rbf kernel works in performing classification for multiclass.

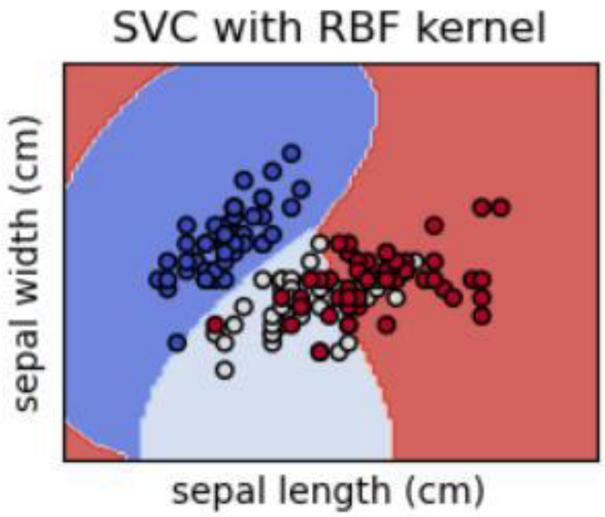

Figure 1.4 Support vector machine classifier with rbf kernel [28]

When training a SVM model with the Radial Basis Function (RBF) kernel, two parameters must be considered: C and gamma. Gamma defines how much influence a single training example has. The larger gamma is, the closer other examples must be to be affected .

2. Random Forest Classifier using Entropy Criterion/Index

The Random Forest algorithm can use the Gini or Entropy criteria in its application. The Gini criterion measures the frequency with which each element of the data set will be mislabeled when randomly labeled. Meanwhile, the Entropy criterion focuses on a measure of information that shows the irregularity of the feature with the target. From the comparison results, the Entropy criterion has slightly better results than the Gini criterion [29].

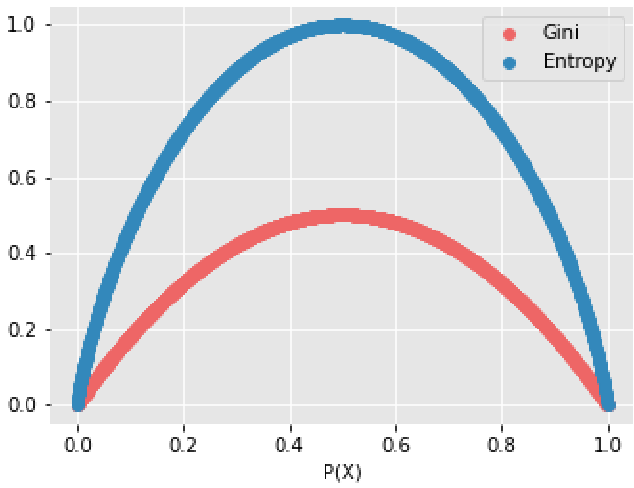

Figure 1.5. Comparison of gini and entropy interval value [29]

But in terms of time duration, the Gini criterion has a much faster duration than the Entropy criterion because the computation process of the Gini criterion is easier than the Entropy criterion which uses logarithmic computation [29].

3. Artificial Neural Network (ANN)

Artificial Neural Network (ANN) algorithms were developed from one of the studies of biology, where Neural Network (NN) plays an important role in the human brain. ANN is an intelligent computing technique that has sustainability inspired by biological neurons [10]. ANN algorithms are massively parallel computing systems consisting of many simple processors with many interconnections. Instead of following a set of laws defined by human experts, ANNs learn the basic laws of the set given a symbolic situation in an example. They are organized in three or more layers, (i.e., input layer, multiple hidden layers, and output layer) [10]. In addition, the analysis activity of these ANNs comes from the relationship between the network's processing units. ANN algorithms are widely used in many fields of study due to their ability to learn from given samples. In addition, ANN algorithms compared to other traditional ML algorithms have advantages in handling random data, fuzzy data, and nonlinear data. Compared to other Neural Network algorithms such as Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN), the ANN algorithm is most suitable for tabular and text data [10].

1.7 Research Methodology

The research methodology can be clarified through Figure 1.6.

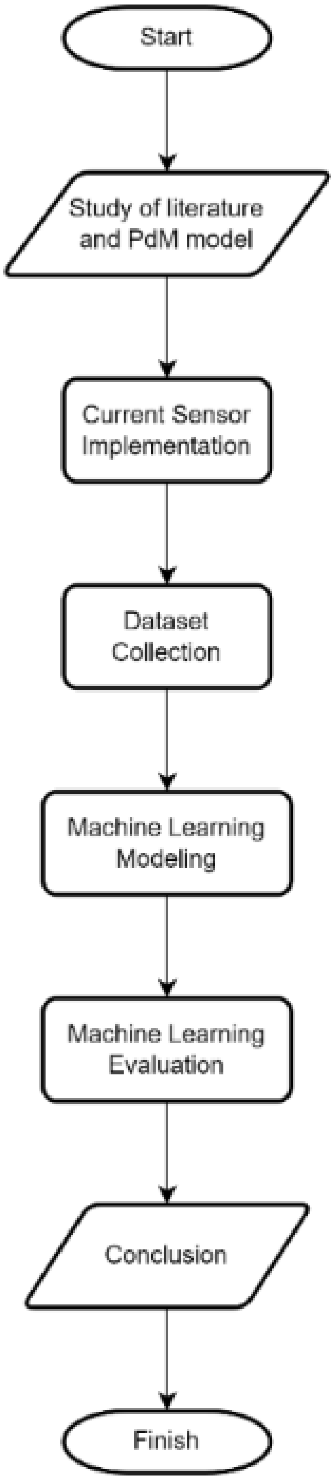

Figure 1.6 Research methodology

1. Study of literature and PdM model

The identification of related problems is carried out through literature studies taken from the latest research results such as journal or conference papers and books related to the topic. In addition, at this stage, system modelling is also carried out which will be the scope of research and problem formulation related to this research. The system is modeled by prototyping a PdM system on RPM equipment with the implementation of current sensors and ML algorithms.

2. Implementation of current sensor using embedded system

The choice of using a current sensor as a monitoring tool for railway point machine (RPM) equipment in this study because the current sensor can be installed on the indoor panel in the station equipment room (ER) so that it is safer, more practical, and more reliable [11]. The implementation of current sensor using embedded systems in the form of NodeMCU ESP8266 modules and Hall Current Transducer.

3. Data collection of equipment conditions

When the RPM equipment is operated by the operator, the working current will be monitored and recorded by the current sensor. Data collection is done by collecting these current sensors reading into csv file. These data will be labeled by three RPM equipment condition classification, including (1) normal situations, (2) warning conditions, and (3) failure conditions. This data will be obtained when the RPM equipment operated in the test-bench at workshop center of the PT KAI (Persero).

4. Modelling the machine learning (ML)

The dataset that has been obtained will be processed in feature extraction to convert the time-series data into some features that relevant to RPM condition classifications. These features and labels will become training and testing dataset for ML modelling. The dataset obtained will be divided into 80% as training data and 20% as testing data. This study will focus on using three ML algorithms, Support Vector Machine (SVM) algorithm, Random Forest (RF) and Artificial Neural Network (ANN) which are the three ML algorithms with the most use for PdM system applications [10].

5. Evaluating the machine learning (ML) models

The performance of three ML models that have been created will be evaluated by using confusion matrix method [24]. The evaluation will be done by comparing the true class and predicted class of testing data. We will find the optimal ML model with the best accuracy from this evaluation.

6. Conclusion

The data analyzed from the evaluation results will be used to answer the research questions.