Comparative Analysis of ARIMA and LSTM Models for Predicting Physical Fatigue in Bandung Workers

Kiki Dwi Prasetyo*, Rifki Wijaya, Gia Septiana Wulandari

School of Informatics, Informatics, Telkom University, Bandung, Indonesia

Email: 1kikidwi@student.telkomuniversity.ac.id, 2rifkiwijaya@telkomuniversity.ac.id, 3giasptiana@telkomuniversity.ac.id

Abstract—In today’s era of rapid economic growth, there is an increasing demand for workers to increase productivity by working longer and harder. However, these demands often lead to irregular and excessive working hours, which can potentially lead to negative consequences, such as physical fatigue—a state in which the body feels tired after physical activity. Factors that influence this fatigue include age, gender, health conditions, workload and work environment. Physical fatigue poses a significant challenge in ensuring workplace safety, especially in the transportation and industrial sectors, as it can reduce overall performance, productivity and quality of work. In addition, physical fatigue also increases the likelihood of decision-making errors and workplace accidents. Predicting physical fatigue is crucial to addressing these challenges. Heart rate serves as a parameter to measure fatigue, given its proven efficacy as a marker to predict physical fatigue, which is derived from the electrocardiogram and regulated by the autonomic nervous system. This research utilizes two machine learning algorithms - ARIMA and LSTM - with heart rate (bpm) and number of steps as variables. Performance evaluation, using Mean Squared Error (MSE) and Root Mean Squared Error (RMSE), showed that the LSTM model outperformed the ARIMA model. The LSTM model showed better performance, with MSE of 0.1108 and RMSE of 0.3329, compared to the ARIMA model with MSE of 0.2397 and RMSE of 0.4895.

Keywords: Autoregressive Integrated Moving Average; Long Short Term Memory; Heart Rate; Time Series; Physical Fatigue

1. INTRODUCTION

The modern era of rapid economic development requires workers to work harder and longer to increase productivity. This demand often results in excessive and irregular working hours. One of the negative impacts that can arise is physical fatigue, which is a condition where the body feels tired after doing physical activity. Fatigue levels can be affected by a range of factors, including age, gender, health conditions, as well as workload and the environment in which one works [1].

In carrying out work, physical fatigue is a serious problem in the field of occupational safety, especially in the transportation and industrial sectors. Physical fatigue can lead to reduced levels of performance, productivity and quality of work [2]. In addition, physical fatigue can also increase the risk of errors in decision-making, and can even lead to accidents in the workplace [3]. Short-term effects of physical fatigue include discomfort, decreased ability and strength, reduced levels of alertness, poor and slowed perception, and feelings of lethargy, drowsiness, and dizziness [1], [2], [4]. This condition of fatigue can also lead to difficulty concentrating, reduced levels of alertness, poor perception, and loss of control over one’s strength [1], [2]. Meanwhile, the long-term effects of physical fatigue involve musculoskeletal and cardiovascular disorders, and can even develop into chronic fatigue syndrome [5]. It’s important to remember that physical fatigue not only affects work performance, but can also have an impact on overall health.

Knowing when physical fatigue will occur is an important step in preventing its negative effects. The way to know when physical fatigue is likely to occur is through prediction. Physical fatigue prediction is the process of estimating the physical fatigue that will occur at a later time by considering current information and the factors that influence it [6]. One of the parameters that can be used to predict physical fatigue is heart rate [7]. In daily life, the most commonly felt change in physiological parameters is heart rate. Heart rate refers to the frequency of heartbeats in one minute. Heart rate can be used to predict physical fatigue because heart rate, which is autonomic nervous system regulates and which is obtained from the electrocardiogram, can be shown to be a potential indicator for forecasting physical fatigue [7]. Heart rate data is time series data, which means that this information is obtained continuously over a certain period of time. Therefore, physical fatigue analysis can be more effective if it considers the temporal aspect of heart rate data, allowing to identify patterns or trends that may be potential predictors of future physical fatigue.

This research aims to predict physical fatigue based on heart rate using two algorithms, namely Autoregressive Integrated Moving Average (ARIMA) and Long Short-Term Memory (LSTM). Both algorithms were chosen because they can learn temporal patterns contained in the data and are suitable for predicting time series data, especially if the prediction time is long [8]. The accuracy of the prediction results from these two algorithms will be compared to evaluate their performance. As a foundation, several previous studies have shown the success of ARIMA and LSTM in predicting time series data, including heart rate data prediction.

As examples, previous research [9] focused on Bitcoin price prediction using ARIMA, GRU, and LSTM models. The results showed that ARIMA gave the best results with MAPE of 2.76% and RMSE of 302.53. Another research [8], which compared several machine learning models such as ARIMA, linear regression, SVM, KNN
regressor, decision tree regressor, random forest regressor, and LSTM, showed that ARIMA, with walk-forward validation and linear regression, outperformed other models in predicting heart rate values in the next time. The findings of these previous studies consistently confirm the reliability of the ARIMA model in handling time series data prediction problems.

Furthermore, research [10], showed that LSTM was able to accurately predict driver fatigue levels using heart rate and percentage of eyelid closure (PERCLOS) data. The prediction results from the LSTM achieved a true positive rate of 75% and an overall accuracy of 88%, demonstrating the ability of the LSTM model to predict fatigue levels with a high degree of accuracy. Then, in [11], a comparison was made between the Long Short-Term Memory (LSTM) and Autoregressive Integrated Moving Average (ARIMA) models for predicting time series data. This paper uses intraday data from the Tehran Stock Exchange (TSE) for 10 years to forecast the next 2 months. LSTM produces better output compared to the ARIMA model. The results show that, although in long-term prediction, the prediction error rates of both models decrease, LSTM significantly outperforms ARIMA in terms of prediction error rates. These studies demonstrate the ability of the LSTM model to predict fatigue levels and time series data with good accuracy.

In performing the comparison, the performance evaluation metrics used are Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). MSE measures the average of the squares of the difference between the predicted value and the true value, while RMSE gives an indication of the prediction error in the same units as the predicted variable. Using these two metrics, the research aims to identify algorithms that provide heart rate-based physical fatigue predictions with lower error rates, thus providing a better understanding of the performance of ARIMA and LSTM in this research.

2. RESEARCH METHODOLOGY

2.1 Research Stages

In this research, a system for predicting physical fatigue is created using two machine learning techniques, namely Autoregressive Integrated Moving Average (ARIMA) and Long Short-Term Memory (LSTM) which will be compared. The flowchart of the system is described in Figure 1 which illustrates the stages to be gone through. The process starts by involving the input dataset which then goes through a preprocessing stage. The dataset is divided into two parts, where one part is used as training data to train the model, while the other part is used as testing data to evaluate the performance of the model. The final process is characterized by an assessment of the system's performance.

2.2 Data collection

This research uses a smartwatch to collect heart rate data for workers in Bandung City for 2 days per worker. Heart rate is considered a physiological parameter that is often felt in everyday life. The variability of heart measurements
is considered an indicator of the level of fatigue. Heart Rate is widely used in identifying fatigue states and measuring performance responses to workload, serving as a direct regulator of cardiovascular function in the autonomic nervous system (ANS). The aim of this research was to analyze heart rate to identify patterns associated with levels of physical fatigue. After data collection for 2 days, interviews were conducted to obtain further information about fatigue time.

2.3 Data preprocessing

Preprocessing refers to a series of steps or processes applied to data before it is utilized for model training or analysis. In this research, preprocessing is carried out with the aim of preparing the data to suit the needs of the model or analysis to be run. In this research, several preprocessing steps were conducted, and they include:

a. Data preparing

The data collected from the smartwatch needs to follow an export step before it can be processed further. The exported data from the smartwatch can be found in the global export data folder, where all exported data will reside. Initially, the data is presented in JSON format, and then needs to be converted to CSV format for further processing. In Table 1, an example of raw data that has been converted to CSV format is shown.

Table 1. Sample Raw Data

<table>
<thead>
<tr>
<th>dateTime</th>
<th>value/bpm</th>
<th>value/confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>09/23/23 00:00:02</td>
<td>99</td>
<td>1</td>
</tr>
<tr>
<td>09/23/23 00:00:07</td>
<td>98</td>
<td>1</td>
</tr>
<tr>
<td>09/23/23 00:00:12</td>
<td>96</td>
<td>1</td>
</tr>
<tr>
<td>09/23/23 00:00:17</td>
<td>95</td>
<td>2</td>
</tr>
<tr>
<td>09/23/23 00:00:32</td>
<td>95</td>
<td>2</td>
</tr>
<tr>
<td>09/23/23 00:00:37</td>
<td>95</td>
<td>2</td>
</tr>
</tbody>
</table>

b. Data interpolation

The quality of the dataset is still not optimal due to several cases where data was not recorded within the 5-second interval, resulting in empty values. Therefore, it is necessary to interpolate the data to fill in the gaps of heart rate information at unrecorded times. Table 2 shows that the data has not gone through the interpolation process, where the data is still not optimal with some time ranges that do not have heart rate data. Meanwhile, Table 3 shows the data that has gone through the interpolation process, where the dataset becomes more optimal with the empty values that have been filled.

Table 2. Before Interpolation Data

<table>
<thead>
<tr>
<th>dateTime</th>
<th>bpm</th>
</tr>
</thead>
<tbody>
<tr>
<td>2023-10-06 00:00:02</td>
<td>57</td>
</tr>
<tr>
<td>2023-10-06 00:00:17</td>
<td>56</td>
</tr>
<tr>
<td>2023-10-06 00:00:22</td>
<td>57</td>
</tr>
<tr>
<td>2023-10-06 00:00:32</td>
<td>58</td>
</tr>
<tr>
<td>2023-10-06 00:00:37</td>
<td>59</td>
</tr>
</tbody>
</table>

Table 3. After Interpolation Data

<table>
<thead>
<tr>
<th>dateTime</th>
<th>bpm</th>
</tr>
</thead>
<tbody>
<tr>
<td>2023-10-06 00:00:02</td>
<td>57</td>
</tr>
<tr>
<td>2023-10-06 00:00:07</td>
<td>56</td>
</tr>
<tr>
<td>2023-10-06 00:00:12</td>
<td>56</td>
</tr>
<tr>
<td>2023-10-06 00:00:17</td>
<td>56</td>
</tr>
<tr>
<td>2023-10-06 00:00:22</td>
<td>57</td>
</tr>
</tbody>
</table>

c. Data labeling

Labeling is done by providing information about heart rate and step data taken using a smartwatch. The label given is "yes" to indicate that the worker is experiencing fatigue and "no" to indicate that the worker is not experiencing fatigue. This labeling process is adjusted to the results of the interviews that have been conducted. Table 4 shows the data that has gone through the labeling process where a fatigue label column is added to indicate the state of fatigue or not on each data entry.

Table 4. Data labelling fatigueness with bpm and step as feature.

<table>
<thead>
<tr>
<th>id</th>
<th>dateTime</th>
<th>bpm</th>
<th>step</th>
<th>fatigue</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID001</td>
<td>2023-09-24 07:22:17</td>
<td>86</td>
<td>2</td>
<td>yes</td>
</tr>
<tr>
<td>ID001</td>
<td>2023-09-24 17:17:07</td>
<td>77</td>
<td>0</td>
<td>yes</td>
</tr>
<tr>
<td>ID002</td>
<td>2023-09-27 06:39:52</td>
<td>83</td>
<td>0</td>
<td>no</td>
</tr>
<tr>
<td>ID002</td>
<td>2023-09-27 19:50:07</td>
<td>63</td>
<td>0</td>
<td>yes</td>
</tr>
</tbody>
</table>
Data normalization is the process of transforming the values in a dataset into a specific range to ensure that all features are similarly scaled [12]. Normalization is performed using methods such as Min-Max Scaling, which converts each value in the dataset into a value between 0 and 1. The normalization formula can be seen in formula (1).

\[ x_{\text{norm}} = \frac{x - \min(x)}{\max(x) - \min(x)} \]  

Figure 2 shows the normalized dataset, where the dateTime, bpm, and step data have changed to a single range of values between 0 and 1.

### 2.4 Split dataset
The process of splitting the dataset is done by separating the dataset into two subsets, which are the training set and the test set. The training subset is used to train the model, while the test subset is used to test the performance of the trained model on data that has never been seen before. The dataset division ratio applied is 70% for the training set and 30% for the test set.

### 2.5 Autoregressive Integrated Moving Average
ARIMA, introduced by Box and Jenkins in 1970, is an efficient tool in forecasting future values in time series [13]. This method is used to describe time series based on observed values and is often applied for short-term forecasting with the advantage of outperforming complex structural models [14]. In the general formula of ARIMA (p, d, q), p refers to the Autoregressive (AR) order which indicates the impact of the dependent variable in the previous period, d is the difference order, and q is the moving average (MA) order which indicates that the independent variable is the residual value in the previous period [15]. This model not only focuses on the influence of past values, but also takes into account the difference and moving average effects, providing a comprehensive view of the time series. ARIMA is widely applied and effective in forecasting periodic series with patterns without random white noise, without seasonality, and can model changes in data efficiently [13]. In an ARIMA model, the forthcoming value of a variable is a linear combination of its previous value and the associated error, represented as follows:

\[ Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \]  

Yt represents the actual value, \( \varepsilon_t \) is the random error at time t, \( \phi_i \) and \( \theta_j \) denote coefficients, and p and q are integers commonly known as autoregressive and moving average parameters, respectively [12].

In this research, the ARIMA method is implemented using the pmdarima library and the auto_arima function. This approach provides higher automatization in determining the best parameters for the ARIMA model. The ARIMA model, which stands for AutoRegressive Integrated Moving Average, is a statistical model that is generally used to analyze time series. The ARIMA model formula involves parameters p, d, and q, which respectively indicate the autoregressive order, differential level, and moving average order. The auto_arima function takes a time series (train_set_arima[‘fatigue’]) as input and automatically carries out the process of finding the best parameters. Additionally, exogenous variables (train_set_arima[columns_to_normalize]) can be included to train the model using additional information. The results of the modeling process are stored in the model_arima variable which is then called to make predictions regarding the level of physical fatigue.

### 2.6 Long Short-Term Memory
Long Short-Term Memory or commonly abbreviated as "LSTM" is a variant of the Recurrent Neural Network (RNN) model. The LSTM model is a type of artificial neural system specifically designed to overcome the limitations of traditional RNNs by using a unique memory cell to store information over long periods. This allows LSTM to effectively handle long-term dependencies and is widely used in various applications, including natural language processing and time series forecasting.
constraints of bursting or disappearing gradients that often occur in long-term dependency learning, especially when there are long enough time intervals [16]. Like any other artificial neural network, LSTM has several hidden layers called memory cells [17], [18]. A memory cell consists of an input gate, forget gate, and output gate [16], [19]. LSTMs are known to be very effective in classification, processing, and forecasting tasks of time series data, especially since there is often an unknown time lag between important events in the time series [20].

![LSTM architecture](image)

**Figure 3. LSTM architecture** [16]

In this research, the Long Short-Term Memory (LSTM) method is implemented with a configuration using an LSTM layer which has a total of 90 neuron units, followed by a Dropout layer with a dropout rate of 0.2. The Dropout layer aims to reduce the potential for overfitting by randomly ignoring a number of neurons during the training process. Next, a Dense layer is added as an output layer with one neuron and a linear activation function.

Once the model structure is determined, the model is compiled using the Adam optimizer, Mean Squared Error (MSE) loss function, and accuracy measures for model performance. The training process begins by configuring the EarlyStopping callback, which incorporates losses on validation data and stops training if no improvement occurs after a specified number of epochs and recovers the best weights.

The model is then drilled using training data with input X_train which includes datetime data, heart measurements, and steps, while y_train contains fatigue label data. Model training was carried out for 50 epochs, with a batch size of 128. Model evaluation was carried out using data validation. With this approach, LSTM models can recognize and utilize information from long-term time series, producing accurate predictions.

2.7 Performance evaluation

The final step in evaluating the performance of ARIMA and LSTM models involves measuring prediction accuracy using Mean Squared Error (MSE). This research chose MSE as the evaluation metric, which is a metric in statistics to calculate the average squared error among the predicted value and the actual value [21]. The Mean Squared Error (MSE) offers a comprehensive measure of proximity between the predicted and actual values. The smaller the MSE value, the higher the accuracy of the model [21]. The prediction error rate is measured as reflected in equation (3) and equation (4).

\[
\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2
\]

\[
\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \hat{x}_i)^2}
\]

3. RESULT AND DISCUSSION

3.1 Result Datasets

The data used in this research includes dateTime, heart rate, and step count information collected through the use of a smartwatch. The source of data came from six workers in the city of Bandung, who provided data for two days for each individual. After the data collection period, interviews were conducted to obtain additional information regarding the time to fatigue. The data obtained from the smartwatches and the interview information were combined in the dataset labeling process. In this process, a "yes" label is used to indicate a fatigued state, while a "no" label indicates a normal or non-fatigued state. Table 5 below shows the results of the preprocessing stage of the dataset:

<table>
<thead>
<tr>
<th>id</th>
<th>date/Time</th>
<th>bpm</th>
<th>step</th>
<th>fatigue</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID001</td>
<td>2023-09-23 01:06:07</td>
<td>97</td>
<td>1</td>
<td>no</td>
</tr>
</tbody>
</table>

**Table 5. Sample datasets**
The data description in Figure 4 reflects a dataset with 149,796 entries and five main columns. First, the 'id' column uses an object data type (string) to provide the numbering of each distinct subject in the dataset. Then, the 'dateTime' column provides timestamp information for each heart rate and step measurement, represented in the int64 (integer) data type. Next, the 'bpm' and 'step' columns store the heart rate (beats per minute) and step count values at each measurement time, both using the int64 (integer) data type. Finally, the 'fatigue' column uses the object data type (string) to present data in the form of text that provides information about fatigue.

The data description in Figure 4 reflects a dataset with 149,796 entries and five main columns. First, the 'id' column uses an object data type (string) to provide the numbering of each distinct subject in the dataset. Then, the 'dateTime' column provides timestamp information for each heart rate and step measurement, represented in the int64 (integer) data type. Next, the 'bpm' and 'step' columns store the heart rate (beats per minute) and step count values at each measurement time, both using the int64 (integer) data type. Finally, the 'fatigue' column uses the object data type (string) to present data in the form of text that provides information about fatigue.

The search process is performed through the use of the auto_arima function. The trace parameter setting was set to True to allow observation of additional information during the search, such as the AIC (Akaike Information Criterion) value for each model tested. This step helps in the selection of the best model based on such evaluation criteria. The result of the search process, the best ARIMA model, can be used to perform physical fatigue level prediction based on the heart rate dataset.

After conducting the search process, it was found that the best ARIMA model to predict the level of physical fatigue is ARIMA (1, 0, 0). This model was then implemented to make predictions based on the heart rate dataset in the test data. The total time taken to fit this model was 112,100 seconds. Next, the best model was tested on the
test data, and the performance evaluation results showed a Mean Squared Error (MSE) of 0.2396575563597266, and a Root Mean Squared Error (RMSE) of 0.48954831900025214.

These evaluation results give an idea of how well the model can generalize the prediction to new data, measuring how close the prediction is to the true value on the test data set. With relatively low MSE and RMSE values, it can be concluded that the ARIMA (1, 0, 0) model is able to provide a fairly accurate prediction of the level of physical fatigue based on the heart rate dataset in the test data.

3.3 Result of Long Short-Term Memory

The results of training and testing for the Long Short-Term Memory (LSTM) model will be illustrated using graphs that can be seen below, there are 2 graphs, namely graphs that describe the change in loss per epoch during the training process and graphs that describe the increase in accuracy during the training process.

**Figure 6. Model Loss per Epoch**

Analysis based on the model loss per epoch graph shows an increase in model accuracy as the training time progresses. The loss on the training data shows a consistent decrease as the number of epochs increases. The final Mean Squared Error (MSE) value reaches 0.11079745451460128, indicating that the model has successfully optimized its performance on the training data. The final Root Mean Squared Error (RMSE) value is 0.3328625159350348. This value indicates that the LSTM model is able to provide predictions that are closer to the true value on the training dataset. This result shows that the LSTM model has learned well from the training data and is able to minimize the prediction error.

Furthermore, the evaluation of training and testing accuracy levels can be seen in the following graph. This graph will provide a clearer picture of how the model performs during the training and testing process. By comparing the accuracy rate between training and testing data.

**Figure 7. Training and Validation Accuracy per Epoch**

The accuracy graph visualizes the performance of the model on the training and validation sets of each epoch. The increase in training accuracy over time reflects the improvement of the model's precision during the training process. There is a positive trend in the accuracy graph, indicating that the model continues to improve in understanding the patterns in the data. Overall, the model performed well, achieving an accuracy rate of 84.33%.

3.4 Comparison of ARIMA and LSTM

The comparison is done by comparing the performance evaluation results of the Autoregressive Integrated Moving Average (ARIMA) and Long Short-Term Memory (LSTM) models. Model performance evaluation is performed using the Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) metrics. The results of the performance evaluation of the two models can be found in Table 4.
From the comparison results table, it can be seen that the LSTM model has better performance than the ARIMA model. This can be observed from the lower Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) values in the LSTM model (MSE=0.1108, RMSE=0.3329) compared to the ARIMA model (MSE=0.2397, RMSE=0.4895). The lower MSE and RMSE values indicate that the LSTM model provides more accurate predictions and is closer to the true value than the ARIMA model.

### 4. CONCLUSION

This research conducted a comparison between two models, namely Autoregressive Integrated Moving Average (ARIMA) and Long Short-Term Memory (LSTM), in estimating the level of physical fatigue based on heart measurement data. From the performance evaluation results, it can be concluded that the LSTM model shows superior performance compared to the ARIMA model. First of all, the LSTM model shows a Mean Squared Error (MSE) value of 0.1108 and a Root Mean Squared Error (RMSE) of 0.3329. On the other hand, the ARIMA model shows an MSE value of 0.2397 and an RMSE of 0.4895. These results indicate that the LSTM model provides predictions that are more accurate and closer to the actual values, with significantly lower MSE and RMSE values compared to the ARIMA model. Further analysis suggests that the superiority of the LSTM model may stem from its ability to capture complex patterns in cardiac measurement time data. LSTM can effectively model and predict the variability of physical fatigue levels based on heart rate signals, which may be difficult to capture by simpler ARIMA models. Thus, these findings support the use of LSTM models as a more effective choice in this context, especially when dealing with complex temporal data.

### ACKNOWLEDGMENT

The author would like to express his deepest gratitude to Telkom University for their invaluable support and facilitation during the implementation of this research. Their commitment and assistance played an important role in the successful implementation of this research. The authors highly appreciate their contributions and collaborative efforts, which have enriched the research process significantly.

### REFERENCES


