INTRODUCTION

Indonesia is the largest archipelago in Southeast Asia with a total area of 7,810,000 Kilometers. Indonesia is a maritime country because the sea area is greater than the land. Fishermen are people with work activities as fish catchers. This work is full of risks, due to high sea waves and wind speed at sea blocking visibility and can also cause the sinking of fishing boats [1]. Fishermen can sail for 4 - 5 days in one departure [2]. In addition, the impact is also felt by beach tourists and coastal residents [3]. Wave height and wind speed information is required by ports, marine transportation and in the design of coastal structures [4].

Modeling that is often used for time-series data is long short term memory (LSTM). Gunawan and friends conducted research on Sharia Stock Prices using LSTM with the parameters used were Layers, Timestep and Epoch for PT Aneka Tambang, Erajaya Swasembada, Kalbe Farma, Semen Indonesia and Wijaya Karya, each obtaining MAPE values of 2.64; 2.24; 1.51; 1.83 and 2.66 [5].

Dzaki and his friends carried out a comparison between LSTM and RNN by testing parameters, namely Epoch and Batchsize, and found that LSTM outperformed RNN in predicting crypto with the LSTM MAPE for Bitcoin being 5.66 and for Ethereum being 4.58 [6].

Prophet is an algorithm developed by Facebook for businesses. The Prophet algorithm functions for time series forecasting. Apart from being able to act as a forecasting method, the prophet algorithm is also able to
handle outliers [7]. Sharma K with his research entitled "Time Series Forecasting Using FB-Prophet" said that Prophet has better performance for empty data and outlier detection besides that Prophet is also very simple [8].

Outliers are data with deviating values [9], [10]. Outlier data should be removed because it affects the normality of the data [11]. Data indicated as outliers will be replaced with the newton interpolation technique. Interpolation is a technique to fill empty values and this technique is capable of estimating non-linear data [12]. Therefore, a prediction system using LSTM with Prophet as outlier detection is proposed.

MATERIAL AND METHODS

Long-Short Term Memory (LSTM)

Long-Short Term Memory is a part of RNN. LSTM has gates that function to add or remove information. The gates are Forget gate, Input Gate, Cell State and Output gate [13], [14]. The way the gates work in LSTM is as follows:

**Forget Gate**

The function of the gate in Figure. 1 is to make a decision that information should be discarded or not in processing.

\[
f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)
\]

**Input Gate**

Figure 2 is the Input Gate architecture whose function is to determine the value to be entered into the cell state.

\[
i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)
\]

\[
\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)
\]

**Cell State**

Figure 3 Cell State architecture which is the memory in a layer. This value is obtained from another gate.

\[
C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t
\]

**Output Gate**

Fig. 4 is the Output Gate architecture whose function is to make decisions from the results obtained from inputs and cell states.

\[
o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)
\]

\[
h_t = o_t \cdot \tanh(C_t)
\]

**Prophet**

Prophet was developed by Facebook for business purposes. Prophet is capable of forecasting and can obtain upper and lower bound values so that it is able to detect outliers [7], [15].

\[
Y(t) = g(t) + s(t) + h(t) + \epsilon(t)
\]

Since the values of wave height and wind speed are not values that can be influenced by vacations, each vacation value calculation = 0 [15], [16].

\[
lower p = 100 \times \frac{1 - intervalwidth}{2}
\]

\[
upper p = 100 \times \frac{1 + intervalwidth}{2}
\]
Description:
\[ g(t) = \text{Trend} \]
\[ s(t) = \text{Seasonal} \]
\[ h(t) = \text{Holiday} \]
\[ \varepsilon (t) = \text{Error Value} \]

\[ \text{interval width} = \text{Randomly generated with maximum a posteriori (MAP)} \]

**Newton Interpolation**

Interpolation is a technique for obtaining a blank value from a line of data [12]. The formula for newton interpolation is as follows:

\[ F_n(x) = b_0 + b_1(x - x_0) + \ldots + b_n(x - x_0)(x - x_1) \ldots (x - x_{n-1}) \]

(10)

**Mean Absolute Percentage Error (MAPE)**

MAPE is a form of model evaluation that is useful for measuring how accurate the model is in carrying out the prediction process [17]–[19].

\[ \text{MAPE} = 100 \times \frac{\sum_{i=1}^{n} |A_t - F_t|}{A_t} \]

(11)

Description:
\( n = \text{Amount of Data} \)
\( A_t = \text{Actual Value} \)
\( F_t = \text{Predicted Value} \)

**Data**

The data used is numerical data with two variables, namely sea wave height (meters) and wind speed (knots) in the form of time series data with 12-hour intervals and a total of 2074 data calculated based on wind measuring devices owned by BMKG in Sumenep Regency. The data was obtained from BMKG Class II Tanjung Perak Surabaya with data location points at the following coordinates:

1. (113.7096, -6.7744) North Point,
2. (114.2693, -6.9011) Northeast Point,
3. (114.0796, -7.0332) East Point, and
4. (113.7956, -7.2874) South Point.

**Wave Height Data Visualization**

Fig 5. Wave height data visualization

**Wind Speed Visualization**

Fig 6. Wind speed data visualization

Figure 5 and Figure 6 are visualizations of the data at the four coordinates of the study site. The data distribution shows a seasonal time series pattern. The colors in the figure represent different data plotting at each point such as blue for the north point, yellow for the northeast point, green for the east point and red for the south point.

**System Flowchart**

The flow of the overall system to be implemented is as follows:

![System Flowchart](image)

Fig 7. System flowchart

The explanation in Figure 7 starts with a start, then the data will be inputted for the data preprocessing process, then the data will be predicted with the best model and obtain the predicted value. Then the calculation of the error value will be calculated using MAPE.

**Flowchart Without Outlier Detection**

![Flowchart Without Outlier Detection](image)

Fig 8. Flowchart without outlier detection
The flow of Figure 8 is model training without outliers, so the data preprocessing is not done at the prophet stage for outlier detection.

**Flowchart With Outlier Detection**

![Flowchart with outlier detection](image)

The flow of Figure 9 in data preprocessing is carried out at the outlier detection stage where when data is detected as an outlier in prophet, the data will be deleted and replaced using the 3rd order Newton interpolation technique. Then the next stage is the training process until obtaining the error value for model evaluation just like Figure 6.

**Test Scenario**

To obtain the model with the smallest error, the test scenario is carried out with the gridsearch method to give each test parameter a chance, the following are the testing stages:

**First Test**

The parameters tested in the first test are:
1. Data Composition (70,30; 80,20; 90,10)
2. Timestep (7, 15, 30)
3. Unit (16, 64, 256)
4. Epoch (10, 100, 200)
5. Learning Rate (0.1; 0.01; 0.001)
6. Batch Size (32, 64, 128, 256)

**Second Test**

This test is conducted to see a good order in the Newton interpolation process. The parameters are the best parameters obtained from the first test and the stages of this test through the outlier detection process are shown in Figure 7. The orders in Newton interpolation tested are orders 3, 5 and 7.

**RESULT AND DISCUSSION**

**Preprocessing Data**

Suppose we have data with the following values:

<table>
<thead>
<tr>
<th>No</th>
<th>Date</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>11/12/2020</td>
<td>1,11017</td>
</tr>
<tr>
<td>2</td>
<td>12/12/2020</td>
<td>0,61264</td>
</tr>
<tr>
<td>3</td>
<td>13/12/2020</td>
<td>0,67832</td>
</tr>
<tr>
<td>4</td>
<td>14/12/2020</td>
<td>0,37062</td>
</tr>
<tr>
<td>5</td>
<td>15/12/2020</td>
<td>0,11973</td>
</tr>
</tbody>
</table>

**Data Normalization**

Data normalization is done to give a number value with a distance of 0 - 1 using the min-max technique. The formula for this technique is:

\[ x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \]  (12)

The data values in Table 1 after going through the formula in equation 12 obtained the following results in Table 2:

**Long Short Term Memory (manual flow)**

In the calculation of LSTM, the initial weight value and bias will be generated randomly between 0 - 1. For example, to calculate LSTM with the first row of normalized data in Table 2 by making the following conditions:

\[ x_t = 1 \] (First row normalized data)
\[ w_f, w_i, w_c, w_o = 0.8 \] (Randomized starting weight)
\[ h_{t-1}, c_{t-1} = 0 \] (Because the First LSTM)
\[ b_f, b_i, b_c, b_o = 1, 0.5, 0.3, 0.1 \]

Using equation 1 to calculate Forget Gate, the following calculation is obtained:

\[ f_t = \sigma(x) = \frac{1}{1 + e^{-x}} \]  (14)
\[ f_t = \frac{1}{1 + 2.718^{-1.3}} \]  (15)
\[ f_t = 0.858126 \]  (16)

Then after calculating the Forget Gate manually in calculations 13 - 16, then calculate the Input Gate using the calculations in equations 2 and 3. The following is an example of the calculation:

\[ i_t = \sigma(x) = \frac{1}{1 + e^{-x}} \]  (17)
\[ i_t = \frac{1}{1 + 2.718^{-1.3}} \]  (18)


\[ i_t = 0.785812 \]
\[ \tilde{C}_t = \tanh (0.8 \ast 0 + 0.8 \ast 1 + 0.3) \]
\[ C_t = \tanh (x) = \frac{2}{1 + e^{-2x}} - 1 \]
\[ \tilde{C}_t = \frac{2}{1 + 2.718^{-2.11}} - 1 \]
\[ C_t = 0.800458 \]

After calculating the calculation on the Input Gate manually in calculations 17 - 23, then calculate the Cell State using equation 4:
\[ C_t = 0.62901 \]

Calculation 24 to obtain the cell state value manually. Then calculate the gate output using equations 5 and 6. Here is an example of the calculation:
\[ a_t = \text{sigmoid}(0.8 \ast 0 + 0.8 \ast 1 + 0.1) \]
\[ a_t = \frac{1}{1 + 2.718^{-0.5}} \]
\[ h_t = 0.10715 \]

For the calculation of the second data, the values of C_t and h_t will use the values obtained from the first calculation, so on until the fifth data and manually use the calculations 13-28 and the results obtained are in Table 3.

<table>
<thead>
<tr>
<th>I</th>
<th>N</th>
<th>f_t</th>
<th>i_t</th>
<th>( \tilde{C}_t )</th>
<th>C_t</th>
<th>a_t</th>
<th>h_t</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.11017</td>
<td>1</td>
<td>0.858126</td>
<td>0.785812</td>
<td>0.800458</td>
<td>0.62901</td>
<td>0.71093</td>
<td>0.10715</td>
<td>0.823864</td>
</tr>
<tr>
<td>0.61264</td>
<td>0.497668</td>
<td>0.815131</td>
<td>0.727852</td>
<td>0.654867</td>
<td>0.989371</td>
<td>0.64194</td>
<td>0.127921</td>
<td>0.755533</td>
</tr>
<tr>
<td>0.67832</td>
<td>0.563982</td>
<td>0.825399</td>
<td>0.74143</td>
<td>0.692859</td>
<td>1.330332</td>
<td>0.657788</td>
<td>0.229602</td>
<td>0.77123</td>
</tr>
<tr>
<td>0.37062</td>
<td>0.253312</td>
<td>0.799983</td>
<td>0.708113</td>
<td>0.595574</td>
<td>1.485976</td>
<td>0.61923</td>
<td>0.178117</td>
<td>0.73304</td>
</tr>
<tr>
<td>0.11973</td>
<td>0</td>
<td>0.758116</td>
<td>0.655302</td>
<td>0.415672</td>
<td>1.398932</td>
<td>0.560322</td>
<td>0.056257</td>
<td>0.674695</td>
</tr>
</tbody>
</table>

Description:
I = Data
N = Normalized Data
D = LSTM Result (Denormalized)

Newton Interpolation (3rd Order)
Suppose the 3rd row data in Table 1 is outlier data, then the data will be deleted and changed to 0 as in Table 4 the data is deleted so that it is without outliers. Then to fill in the 0 data, the 3rd order newton interpolation technique is carried out with the calculation of each order as follows:

<table>
<thead>
<tr>
<th>x</th>
<th>y(F(x_n))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1,1017</td>
</tr>
<tr>
<td>2</td>
<td>0,61264</td>
</tr>
<tr>
<td>3</td>
<td>0,37062</td>
</tr>
<tr>
<td>4</td>
<td>0,11973</td>
</tr>
</tbody>
</table>

1st Order Calculation
\[ F[x_1, x_0] = \frac{F(x_1) - F(x_0)}{(x_1 - x_0)} \]
\[ F[x_1, x_0] = \frac{0.49753 - 1.11017}{(2-1)} = -0.49753 \]

2nd Order Calculation
\[ F[x_2, x_1] = \frac{F(x_2) - F(x_1)}{(x_2 - x_1)} \]
\[ F[x_2, x_1] = \frac{0.37062 - 0.61264}{(3-2)} = -0.24202 \]

3rd Order Calculation
\[ F[x_3, x_2, x_1, x_0] = \frac{F(x_3) - F(x_2)}{(x_3 - x_2)} \]
\[ F[x_3, x_2, x_1, x_0] = \frac{0.11017 - 0.37062}{(4-3)} = -0.25089 \]
\[ F[x_3, x_2, x_1, x_0] = \frac{0.006745 - (-0.004435)}{(4-1)} = 0.0037 \] (40)

Table 5. Table for Newton Interpolation

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>Orde 1</th>
<th>Orde 2</th>
<th>Orde 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1,11017</td>
<td>-0.49753</td>
<td>0.006745</td>
<td>0.0037</td>
</tr>
<tr>
<td>2</td>
<td>0,61264</td>
<td>-0.24202</td>
<td>-0.004435</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0,37062</td>
<td>-0.25089</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0,11973</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

With Table 5, the calculated values of order 1 - 3 calculated manually in calculations 29 - 40 are written down to be compiled and then calculated using equation 10 for the newton interpolation formula. Here is the calculation of newton interpolation:

\[ F(2,5) = 1,11017 + (-0.49753)(2,5 - 1) + 0.006745(2,5 - 1)(2,5 - 2) + 0.0037(2,5 - 1)(2,5 - 2)(2,5 - 3) \] (41)

\[ F(2,5) = 0.36754625 \] (42)

So in Table 4 it is assumed that the center value of x is 2.5 so that the assumption in Table 1 if the 3rd row data is an outlier and is replaced using newton interpolation, the result is 0.36754 resulting from the calculation of 41 and 42 manually.

**First Test Results**

Testing using the gridsearch method makes the computation time quite long, because the first test has many parameters that need to be tested and the data tested has 4 points on 2 test variables, namely wave height and wind speed. After testing 972*8, which is 7,776 times, the parameter combination with the smallest MAPE can be seen in Table 6 and Table 7:

Table 6. First test results wave height

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>90:10</td>
<td>7</td>
<td>256</td>
<td>200</td>
<td>0.01</td>
<td>128</td>
<td>13,32</td>
<td></td>
</tr>
<tr>
<td>TL</td>
<td>90:10</td>
<td>7</td>
<td>64</td>
<td>200</td>
<td>0.1</td>
<td>128</td>
<td>13,32</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>80:20</td>
<td>30</td>
<td>16</td>
<td>200</td>
<td>0.1</td>
<td>128</td>
<td>9,32</td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>80:20</td>
<td>30</td>
<td>16</td>
<td>200</td>
<td>0.1</td>
<td>128</td>
<td>8,85</td>
<td></td>
</tr>
</tbody>
</table>

Table 7. First test results wind speed

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>90:10</td>
<td>7</td>
<td>16</td>
<td>100</td>
<td>0.01</td>
<td>32</td>
<td>14,74</td>
<td></td>
</tr>
<tr>
<td>TL</td>
<td>90:10</td>
<td>30</td>
<td>256</td>
<td>200</td>
<td>0.1</td>
<td>64</td>
<td>14,86</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>90:10</td>
<td>30</td>
<td>16</td>
<td>100</td>
<td>0.001</td>
<td>32</td>
<td>17,11</td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>90:10</td>
<td>30</td>
<td>16</td>
<td>200</td>
<td>0.001</td>
<td>32</td>
<td>14,52</td>
<td></td>
</tr>
</tbody>
</table>

Description:
A = Location
B = Data Composition Value
C = Timestep Value
D = Unit Value
E = Epoch Value
F = Learning Rate Value
G = Batch Size Value
H = MAPE Gain
U = North Point Data
TL = Northeast Point Data
T = East Point Data
S = South Point Data

**Second Test Results**

The outliers detected by Prophet on each data can be seen in Table 8.

Table 8. Number of outliers in the data

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>GA</th>
<th>KA</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>133</td>
<td>117</td>
<td></td>
</tr>
<tr>
<td>TL</td>
<td>123</td>
<td>116</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>100</td>
<td>124</td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>96</td>
<td>125</td>
<td></td>
</tr>
</tbody>
</table>

Description:
GA = Wave Height Data
KA = Wind Speed Data

After the data is detected outlier, the data is replaced with the newton interpolation technique with testing the order performed is 3, 5 and 7 and the value is evaluated with MAPE. The acquisition of MAPE can be seen in Table 9 and Table 10 for each order in the test:

Table 9. Second test results wave height

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>Orde 3</th>
<th>Orde 5</th>
<th>Orde 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>14,75</td>
<td>14,29</td>
<td>16,99</td>
<td></td>
</tr>
<tr>
<td>TL</td>
<td>14,07</td>
<td>19,01</td>
<td>13,56</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>11,29</td>
<td>10,64</td>
<td>10,34</td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>11,05</td>
<td>12,32</td>
<td>9,87</td>
<td></td>
</tr>
</tbody>
</table>

Table 10. Second test results wind speed

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>Orde 3</th>
<th>Orde 5</th>
<th>Orde 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>14,92</td>
<td>18,15</td>
<td>17,99</td>
<td></td>
</tr>
<tr>
<td>TL</td>
<td>14,85</td>
<td>16,57</td>
<td>15,97</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>15,14</td>
<td>18,19</td>
<td>17,25</td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>16,33</td>
<td>15,55</td>
<td>14,59</td>
<td></td>
</tr>
</tbody>
</table>

The overall results of the test provide a decision to use the parameters that obtain the smallest error value (MAPE). The following is a summary of the best MAPE results in Table 11 and Table 12.

Table 11. Best MAPE Results Wave Height

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>Orde 0</th>
<th>Orde 3</th>
<th>Orde 5</th>
<th>Orde 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>13,32</td>
<td>14,75</td>
<td>14,29</td>
<td>16,99</td>
<td></td>
</tr>
<tr>
<td>TL</td>
<td>13,32</td>
<td>14,07</td>
<td>19,01</td>
<td>13,56</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>9,32</td>
<td>11,29</td>
<td>10,64</td>
<td>10,34</td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>8,85</td>
<td>11,05</td>
<td>12,32</td>
<td>9,87</td>
<td></td>
</tr>
</tbody>
</table>
Table 12. Best MAPE results wind speed

<table>
<thead>
<tr>
<th></th>
<th>Orde 0</th>
<th>Orde 3</th>
<th>Orde 5</th>
<th>Orde 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>14.74</td>
<td>14.92</td>
<td>18.15</td>
<td>17.99</td>
</tr>
<tr>
<td>TL</td>
<td>14.86</td>
<td>14.85</td>
<td>16.57</td>
<td>15.97</td>
</tr>
<tr>
<td>T</td>
<td>17.11</td>
<td>15.14</td>
<td>18.19</td>
<td>17.25</td>
</tr>
<tr>
<td>S</td>
<td>14.52</td>
<td>16.33</td>
<td>15.55</td>
<td>14.59</td>
</tr>
</tbody>
</table>

Description:

Orde 0 = data without interpolation (does not go through the outlier detection stage)

The smallest MAPE value is marked with a yellow shell in Tables 11 and 12. Therefore, the following are the results of obtaining the best parameters at each data location point.

**North Wave Height Model**

Table 13 is the LSTM model parameters on the wave height data at the northern location with a MAPE of 13.32. The plotting of actual and predicted data is in Fig. 10 with the red line is the actual data and the blue line is the predicted data:

Table 13. North wave height parameter

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>7</td>
<td>256</td>
<td>200</td>
<td>0.01</td>
<td>128</td>
<td>DA</td>
</tr>
</tbody>
</table>

Fig 10. North wave height model

**Northeast Wave Height Model**

Table 14 is the LSTM model parameters on the wave height data at the northeast location with a MAPE of 13.32. The plotting of actual and predicted data is in Fig. 11 with the red line is the actual data and the blue line is the predicted data:

Table 14. Northeast wave height parameter

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>7</td>
<td>64</td>
<td>200</td>
<td>0.1</td>
<td>128</td>
<td>DA</td>
</tr>
</tbody>
</table>

Fig 11. Northeast wave height model

**East Wave Height Model**

Table 15 is the LSTM model parameters on the wave height data at the eastern location with MAPE 9.32. The plotting of actual and predicted data is in Fig. 12 with the red line is the actual data and the blue line is the predicted data:

Table 15. Parameter Wave Height East

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>80:20</td>
<td>30</td>
<td>16</td>
<td>200</td>
<td>0.1</td>
<td>128</td>
</tr>
</tbody>
</table>

Fig 12. East wave height model

**South Wave Height Model**

Table 16 is the LSTM model parameters on the wave height data at the southern location with a MAPE of 8.85. The plotting of actual and predicted data is in Fig. 13 with the red line is the actual data and the blue line is the predicted data:

Table 16. South wave height parameter

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>80:20</td>
<td>30</td>
<td>16</td>
<td>200</td>
<td>0.1</td>
<td>128</td>
</tr>
</tbody>
</table>

Fig 13. South wave height model

**North Wind Speed Model**

Table 17 is the LSTM model parameters on the wind speed data at the northern location with MAPE 14.74. The plotting of actual and predicted data is in Fig. 14 where the red line is the actual data and the blue line is the predicted data:

Table 17. North wind speed parameter

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>90:10</td>
<td>7</td>
<td>16</td>
<td>100</td>
<td>0.01</td>
<td>32</td>
</tr>
</tbody>
</table>
**Northeast Wind Speed Model**

Table 18 is the LSTM model parameters on wind speed data at the northeastern location with a MAPE of 14.85. The plotting of actual and predicted data is in Fig. 15 with the red line is the actual data and the blue line is the predicted data:

Table 18. Northeast wind speed parameter

<table>
<thead>
<tr>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>90:10</td>
<td>30</td>
<td>256</td>
<td>200</td>
<td>0.01</td>
<td>64</td>
<td>O3</td>
</tr>
</tbody>
</table>

**East Wind Speed Model**

Table 19 is the LSTM model parameters on the wind speed data at the eastern location with a MAPE of 15.14. The plotting of actual and predicted data is in Fig. 16 with the red line is the actual data and the blue line is the predicted data:

Table 19. East wind speed parameter

<table>
<thead>
<tr>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>90:10</td>
<td>30</td>
<td>16</td>
<td>100</td>
<td>0.001</td>
<td>32</td>
<td>O3</td>
</tr>
</tbody>
</table>

**South Wind Speed Model**

Table 20 is the LSTM model parameters on wind speed data at the southern location with a MAPE of 14.52. The plotting of actual and predicted data is in Fig. 17 with the red line is the actual data and the blue line is the predicted data:

Table 20. South Wind Speed Parameter

<table>
<thead>
<tr>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>90:10</td>
<td>30</td>
<td>16</td>
<td>200</td>
<td>0.001</td>
<td>32</td>
<td>DA</td>
</tr>
</tbody>
</table>

Fig 14. North wind speed model

Fig 15. Northeast wind speed model

Fig 16. East Wind Speed Model

Fig 17. South Wind Speed Model

Description:

DA = Original Data (Without outlier detection stage)
O3 = Data with 3rd order interpolation

The following are the results of the test analysis and the model obtained from the test scenario:

1. The split data, timestep, unit, epoch, learning rate and batch size parameters affect the error value as evidenced by the change in the error value generated by each parameter test change.
2. Outliers also affect changes in error values (Affecting data).

Order 3 in newton interpolation is the best order by evaluating using the smallest MAPE value from the test results on the best parameters obtained from the first test, attributes for clustering with a fairly high Silhouette value between 0.979 – 0.811 at MinPts 3 – 5 with the number of clusters = 3. The DBSCAN method is better than OPTICS. While in this study using 17 attributes with a fairly high variance and outlier data. The highest Silhouette score is 0.442 in the multistage clustering method (OPTICS – K-Means) with the number of clusters = 8. This method is better than DBSCAN – K-Means, K-Means, and Agglomerative.
CONCLUSION

In the early stages several transformations After testing using Gridsearch for 7,800 tests, the conclusions that can be drawn are:

1. There are so many LSTM parameter values that the testing process using gridsearch takes a very long time. For data on sea wave height and wind speed, the MAPE value for each data point location is different. The values in parameters such as data composition, Timestep, Unit, Epoch, Learning Rate and Batch size at each location are also different to get the smallest MAPE value. The best MAPE values at each of the north, northeast, east and south locations are ((13.32; 14.74), (13.32; 14.85), (9.32; 15.14), (8.85; 14.52)).

2. Newton interpolation of order 3 is the best order with the lowest error between orders 5 and 7.

REFERENCES


[11] Y. A. Venanda, “Hubungan efikasi diri (self efficacy) dengan prokrastinasi...