ANALYSIS AND PREDICTION OF FACTORS CONCERNING WIND ENERGY USING LSTM

Abstract

A comprehensive analysis and prediction of all feature concerning wind energy is the proposed topic of this project. For the breakdown, analysis, and study of the data, we use a deep learning method based on recurrent neural network (RNN), popularly known as LSTM (Long Short -Term Memory). An accurate prediction of wind energy at a particular point in time is difficult as, the factors affecting the wind are highly erratic and does not follow a trend as we hoped. Thus, for an acceptable prediction we use LSTM as its very useful while dealing with time series prediction. With the help of past data of wind in a certain area for a period, we can predict the wind energy for a point of time in the future, solely relaying on the trend observed in the years of data. Here, it is possible to make prediction for periods of time like a day, or a week. Although the prediction may differ because the of erratic factors, the general prediction can be considered useful for energy production. The dataset used is from NREL (National Renewable Energy Laboratory) in an area in State of Tamil Nādu over a years' time (2007). The database used here contains changes in 5 different features for 8760 rows of data. The data training gave satisfactory results with root mean squared error of 1.456 and a variance of 0.927. The predictions plots for the features were plotted for better understanding.

1.Introduction

Wind energy has been gaining more and more importance in recent times because of its characteristics as a reliable renewable energy source. As this trend continues and the global capacity of wind energy increases, there occurs a need to predict the pattern and features of wind for an efficient and comprehensive wind energy production. Analysis and prediction of wind speed, power generated, wind direction, pressure, and air temperature from the known data in the area will help us in creating suitable conditions for power generation with maximum efficiency and helps in proper scheduling of the working of such a PowerStation. Although there are many erratic factors affecting wind power, the power generation is dependent mainly on wind speed. So, when we carefully analyze wind speed pattern for a certain amount of time, it possible to get a satisfactory prediction.

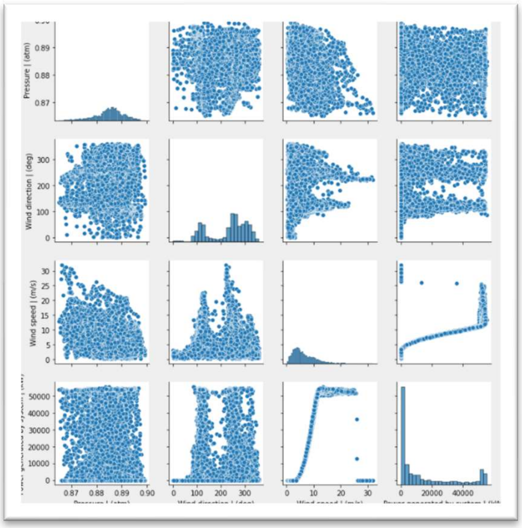
Using this time series pattern, we can gain useful information that can be used practically.

2.Dataset

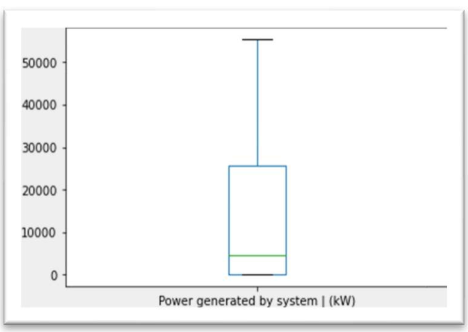
A reliable training and testing data set is an important factor for the proper functioning of the system to give a satisfactory output. Hence, the dataset used needs to carefully selected. The dataset used in this project was derived from the extensive database of National Renewable Energy Laboratory (NREL). The data was recorded during the year of 2007 from the state of Tamil Nādu, with 5 different features observed with respect to time, amounting up to 8760 rows of observation. Features recorded here includes wind speed (m/s), wind direction (deg), pressure (atm), power generated (kw).

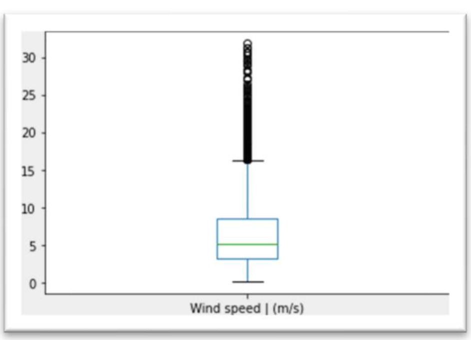
For further analysis of data itself we make use of the seaborn (imported as sns). This helps in identifying the nature of distribution of data, correlation between the features itself present in the data. Here, such plots include are pair plot, boxplot, and heatmap. Other data analyzing function were also used

Pair plot of the data will show as vividly how each feature interacts with each other by plotting them directly. This will help us gain more understanding about the data.



The boxplot of the features points out the measure of distribution of the data for that feature. Data will be catagorically divided to three quartiles as the graph represents factors like minimum,maximum, median etc.. The boxplots plotted here for the power generated, wind speed features.





Other than plots, functions can also help in analyzing the data, to acquire information about the data or to describe the data.

Data info:

```
DatetimeIndex: 8760 entries,
2007-01-01 00:00:00 to 2007-12-31
23:00:00
Data columns (total 4 columns):
 #   Column
Non-Null Count  Dtype
---  ------
--------------  -----
 0   Pressure | (atm)
8760 non-null   float64
 1   Wind direction | (deg)
8760 non-null   int64
 2   Wind speed | (m/s)
8760 non-null   float64
 3   Power generated by system |
(kW)  8760 non-null   float64
```

Data describe:

Pressure | (atm) Wind direction | (deg) Wind speed | (m/s) Power generated by system | (kW)

count 8760.000000 8760.000000 8760.000000 8760.000000

mean 0.884977 228.458219 6.308794 14803.923135

std 0.005786 81.872038 4.221430 18801.398577

min 0.865102 0.000000 0.169000 0.000000

25% 0.882163 137.000000 3.288750 0.000000

50% 0.885612 252.000000 5.239000 4676.360000

75% 0.888562 294.000000 8.537250 25765.400000

max 0.899212 360.000000 31.950000 55485.700000

The two-dimensional graphical representation of data, more commonly known as heatmap. The visualization of heat map using color can be easily used to direct users to areas that are seen to be more important.

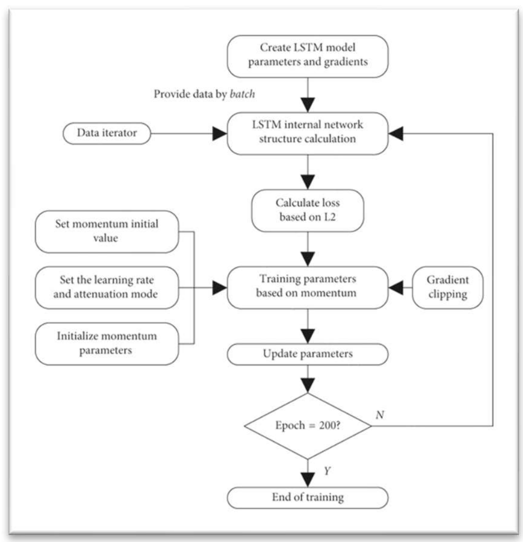
Depicting data as graphs and gaining all related information about the same helps us analyze and predict our desired output from the system.

3.Long Short -Term Memory Model (LSTM)

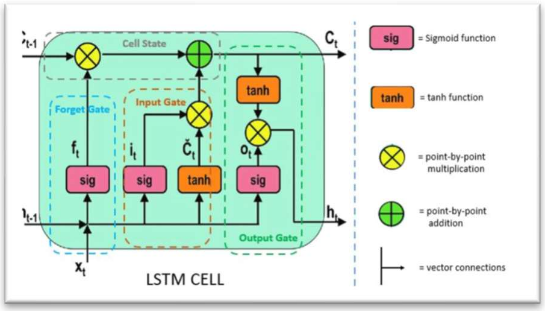
Selecting the learning model needed for the data training is an integral part in visualization and prediction of desired outputs. The most suitable model for training the dataset under consideration is LSTM (Long Short- Term Memory Model).

Being a part of RNN (Recurrent Neural Network), LSTM is structured to remember and predict long term dependencies that are trained with time series data. The advantage of a LSTM model over traditional RNN networks is that the problem of exploding and vanishing gradient in backpropagation, present in conventional RNN is solved.

The LSTM architecture aims to provide a short-term memory for RNN that can last thousands of timesteps, thus "long short-term memory." Training process of the LSTM model is depicted in given flow diagram.



The LSTM training model generally has three gates which includes forget gate, input gate, and output gate. Along with these gates, the cell state is also used for its working. These factors enable the LSTM model to easily learn, unlearn or to store the information selectively. The gates perform function assigned to them, while the cell state act as a horizontal line in which the information flows without any changes.



The components present in the LSTM model are clearly depicted in the above diagram. Among them;

Forget Gate: This gate determines the importance of a data i.e.; which of the data requires consideration and which of the data needs to be rejected. Current input and hidden state information is used to determine the output of the forget gate.

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

Output Gate: This gate is used in determining the next hidden state using the information from previous inputs known. By analyzing the output values of current and previous hidden state, this gate determines which value the next hidden state should carry.

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

$$h_t = o_t * \tanh(C_t)$$

Input Gate: This gate is used in updating cell status mainly by two operations; first to get input gate at t and second to get value generated by tanh. For these also current state and previous hidden state are taken as input variables.

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

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

Cell State: The central role in an LSTM is carried out by this memory cell. The gates regulate and control the information. Adding or removing of the data and flow of information in and out of the cell is also done by the gates. The new cell state derived by the help of previous cell state, forget and input gates.

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

LSTM is used in artificial intelligence and deep learning. This recurrent network is capable of processing entire sequences of data in a self-supervised learning manner. Feedback connections present makes this standout from regular feedforward networks. This is one of most powerful system used in forecasting and greatly helps when there is data with longer trend. It is also considered better than SVM because of its ability to learn or forget data in more efficient manner.

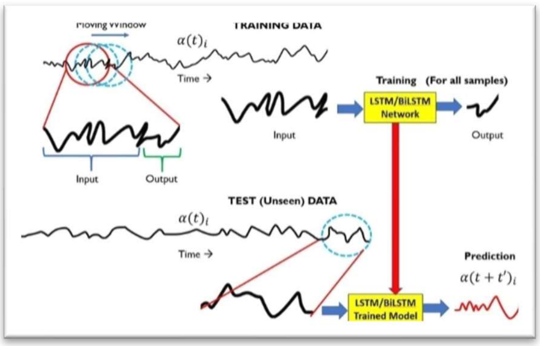
4.Experimentation

4.1 Experimentation Methodology

The dataset that was chosen was pre-processed to create to a suitable data so that we can easily read it into the program and can also navigate through the data with ease. After describing the data with the help of inbuilt function, we now know all necessary attributes of the data which in turn we use later. As the objective of program, we plan to predict the trend of each of the features present in the data which includes power generated, wind speed, pressure etc. The information about the prediction of this data is crucial in determining the working conditions of the wind turbine.

In order to predict each of the features, we first must make sure that when we are running prediction of one data the other feature does not interfere. To make this happen, as a precaution we drop all the rest of features other than the one we are considering from the dataset temporarily. The times series analysis of the data through LSTM training is the next step in the programming methodology.

Since we need to test the prediction after training the network, we need to split the dataset to two parts: One part to train the model by LSTM and other part to test and validate the result. We have 8760 rows of data from which we can create a training set. For our convenience, in this experiment data set has been divided into 60-40 in which 60% of the data is used for training and 40% data is used for testing the prediction. This makes the number of elements in training set to be 5248 and testing set to be 3496.



Next, we train the data. Before the training part, we need to create a specific dataset for the LSTM to traverse through for the time series analysis. So, we create 4 different set in which 2 is for training and 2 for testing. For this we also need to give a timestep or lookback. Timestep/lookback is a common term used in LSTM which signifies the number of steps (apart from the pattern learned) an LSTM model will use to predict the next result. The LSTM model shows three dimensional properties so, we change the dataset to satisfy the LSTM condition. Thus, the dataset is reshaped to three dimensions from two dimension. We only need to reshape the training data as only that needs to go through LSTM training model.

From the keras software we import all the functions needed to run the LSTM model. We add, compile, and fit the training data using LSTM and trains the neural network for 500 epochs. Validation of LSTM is also carried out, the validation test set assesses the ability of the neural network to predict based on new conditions that were not part of the training set. The validation is performed with the last 20% of the data that was separated from the beginning 80% of data. The model was compiled with 'adam' optimizer and loss with mean squared error. After completion, we plot the prediction curves and loss curves to verify the training of the neural network. The efficiency of the neural network is inferred by the root mean square error and variance calculated. We can also observe and change the epochs to get more well-defined curves and to reduce the error in training. The same procedure was carried out for each of the features to get their respective prediction and loss curves.

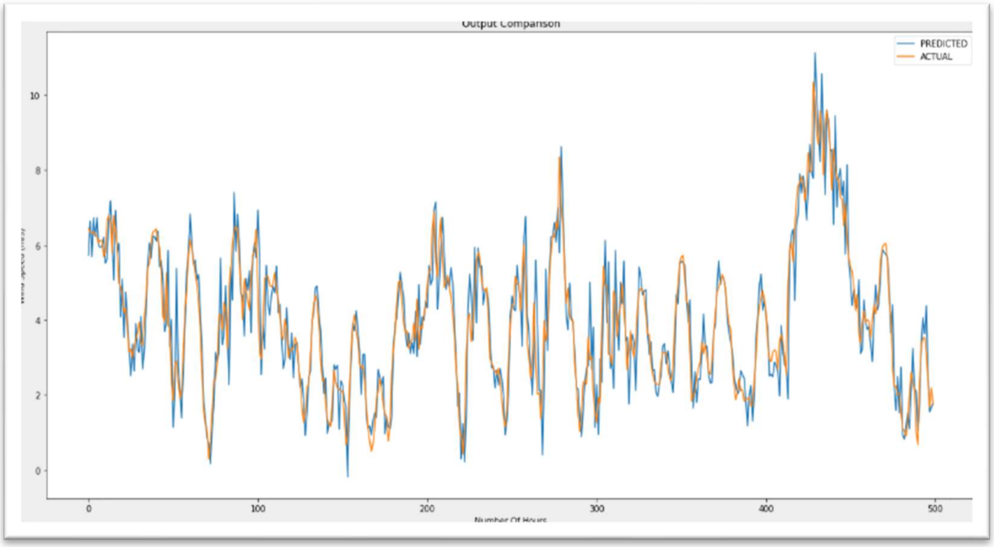
4.2 Experimental Results

The easiest way to understand the prediction accuracy of any neural network is through graph plots. In this experiment, we need to predict all the features accurately in order to apply it in real life if needed. The training done here is a pure time series analysis. We are predicting power generated, wind speed and all other factors without any knowledge of future weather. Here, LSTM analyzes the prior data and tries to get useful information from the trends and patterns of available data. The data training gave satisfactory results with root mean squared error of 1.456 and a variance of 0.927.

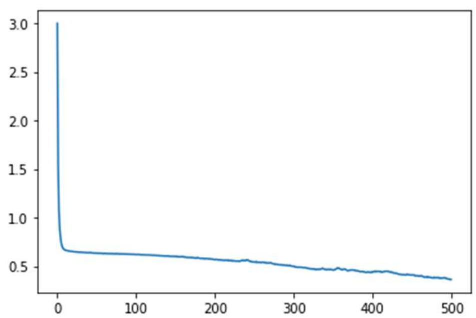
```
Model: "sequential"
___________________________________________
_____________________
 Layer (type)                Output Shape
Param #
===========================================
=====================
 lstm (LSTM)                 (None, 100)
40800

 dense (Dense)               (None, 1)
101

===========================================
=====================
Total params: 40,901
Trainable params: 40,901
Non-trainable params: 0
```

The resulting graphs of features from training is given below:

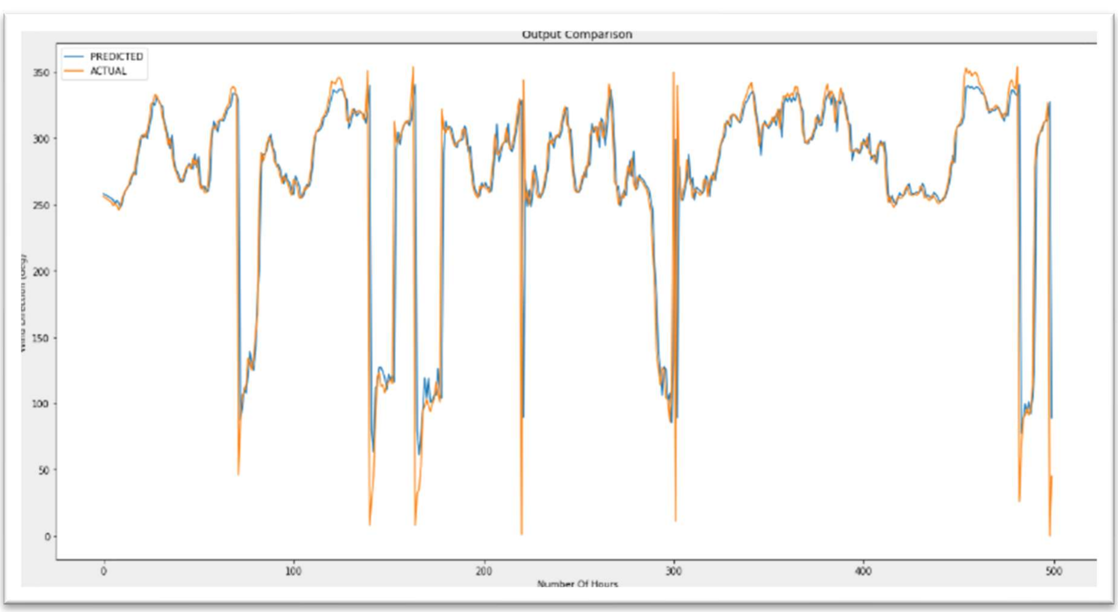
- Wind Speed V/s Number of Hours:



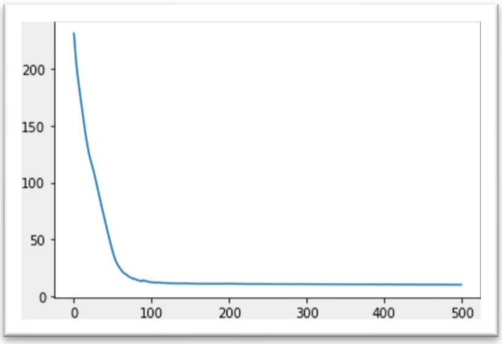
➢ Loss Plot:



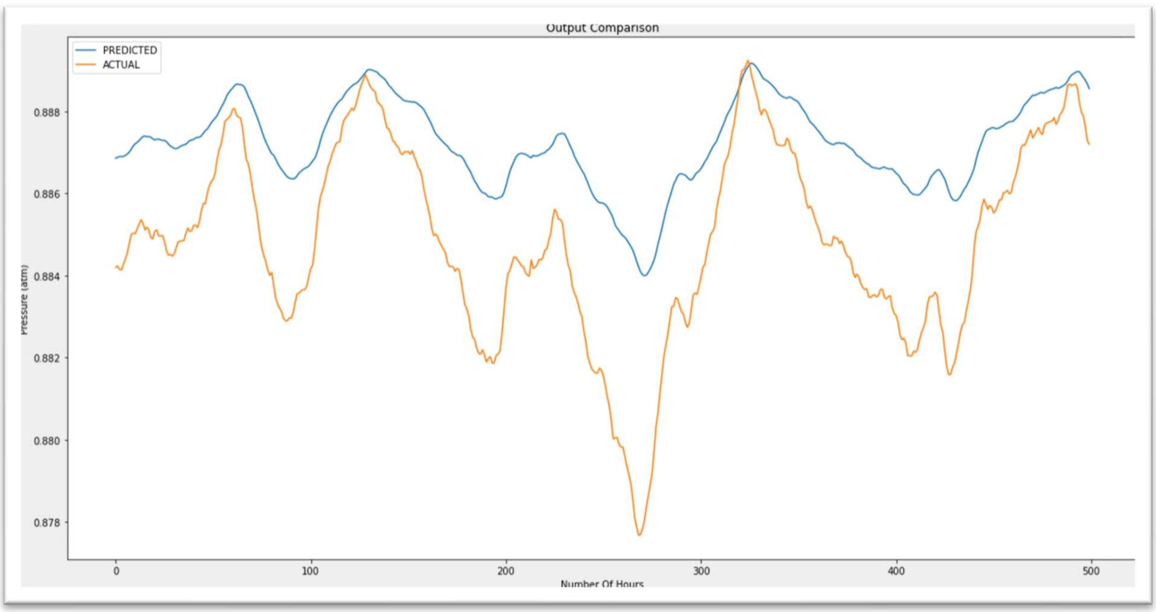
- Wind Direction V/s Number of Hours:



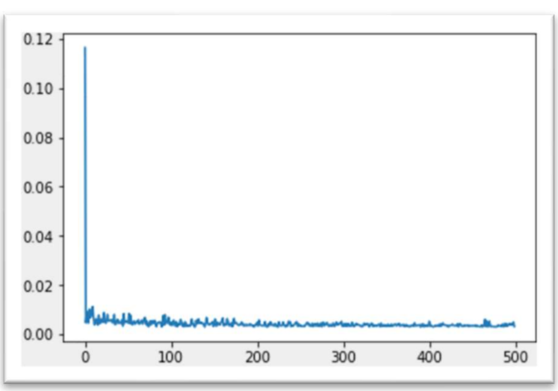
➢ Loss Plot:



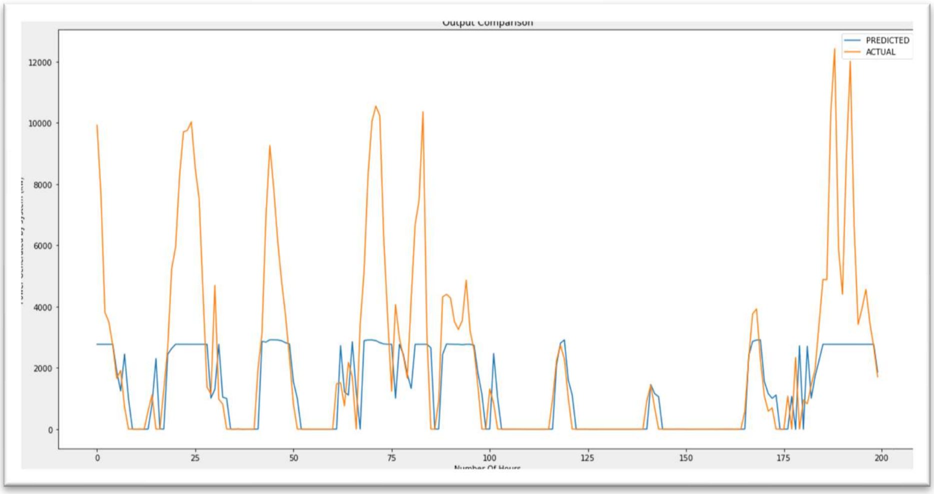
• Pressure V/s Number of Hours:



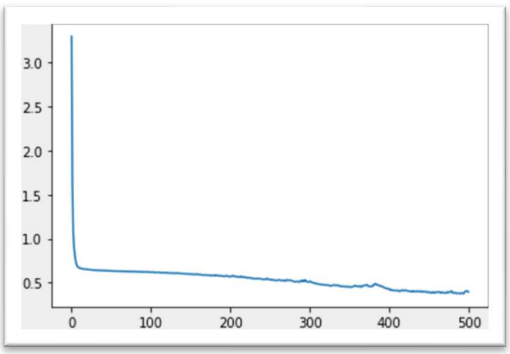
➢ Loss Plot:



- Power Generated V/s Number of Hours:



- Loss Plot:



Code (GitHub): https://github.com/isaac-wq/Wind-Data-Analysis-And-Prediction.git

5. Conclusion

In this experiment, the prediction and analysis of wind data was completed and all output prediction curves were observed. The overall training of the neural network and output plots were satisfactory and a good training of given data was shown. LSTM completed the time series-based prediction completely and proved that it is one of the best models for this type of analysis. But we also observed that if the wind speed is less than 4 m/s the power generated by the system is zero.

LSTM was not able to learn this pattern as this is not the part which it can understand in time series analysis. If we want a better result, we need to create a model with a combination of decision tree and LSTM, which can be very complex. Also, the wind speed prediction may be more erratic and unpredictable if we consider many other real time factors such as unexpected weather changes. This project was done considering that wind will follow a certain pattern over a large period of time.

Thus, the results of this project can be considered feasible to extract useful wind energy without many problems.

6.References:

1. Long short-term memory - Wikipedia

https://en.wikipedia.org/wiki/Long_short-term_memory

2. introduction-to-long-short-term-memory-lstm

https://www.analyticsvidhya.com/blog/2017/12/fundamentals-of-deep-learning-introduction-to-lstm/

3. LSTM Networksapmonitor.com

https://apmonitor.com/do/index.php/Main/LSTMNetwork#:~:text=LSTM%20Prediction%20Validation,the%20beginning%2080%25%20of%20data.

4. What is LSTM - Introduction to Long Short-Term Memory

https://intellipaat.com/blog/what-is-lstm/

5. Introduction to LSTM Units in RNN | Pluralsight

https://www.pluralsight.com/guides/introduction-to-lstm-units-in-rnn

6. LSTM and its equations. LSTM stands for Long Short-Term Memory

https://medium.com/@divyanshu132/lstm-and-its-equations-5ee9246d04af

7. Training process of the LSTM model. | Scientific Diagram

https://www.researchgate.net/figure/Training-process-of-the-LSTM-model_fig3_345450231

8. Time Series - LSTM Model

https://www.tutorialspoint.com/time_series/time_series_lstm_model.htm#:~:text=An%20LSTM%20module%20has%20a,only%20a%20few%20linear%20interactions.

9.Wind Data and Tools | Wind Research | NREL

https://www.nrel.gov/wind/data-tools.html

10. Forecasting Wind Power Generation Using Artificial Neural Network: “Pawan Danawi”—A Case Study

https://www.hindawi.com/journals/jece/2021/5577547/