IBM Hack Challenge 2020

Problem Statement
Predicting the energy output of wind turbine based on weather condition

Presented by

Mahesh Theng

Government College of Engineering, Nagpur

Team Name: WinEner

July 2020

Index

- INTRODUCTION
- LITERATURE SURVEY
- AIMS and OBJECTIVES
- PROPOSED SOLUTION
- THEORITICAL ANALYSIS
- EXPERIMENTAL INVESTIGATIONS
- FLOWCHART
- RESULTS
- CONCLUSION
- FUTURE SCOPE
- BIBILOGRAPHY

INTRODUCTION

- Wind energy plays a major role in providing energy worldwide.
- Renewable energies are set to conquer the global energy system faster than any other fuel in history.
- A wind farm's energy output depends heavily on the weather conditions present at its site.
- A wind performance forecast is an estimation of the production expected of one or more wind turbines.
- If the output can indeed be predicted more effectively, the energy providers can more effectively organize the joint development of various energy sources to avoid expensive overproduction.

LITERATURE SURVEY

- Global energy demand is increasing, and the use of nuclear power, traditional sources of energy such as coal and oil is either considered unsafe or leads to a large amount of CO_2 emission.
- One the other hand wind is a natural free energy source.
- This is extremely unpredictable, which is a major problem for the incorporation of massive wind power into an energy grid.
- Present power production by wind farm is very less than the requirements for solving various problems.
- With the improvement of forecasting wind speed and wind direction, it is possible to maximize power production of a wind farm.

AIMS and OBJECTIVES

- This project facilitates:
  - ✓ Identifying most significant features for wind power prediction.
  - ✓ Continuous learning and model improvement by hybrid ensemble with data and function perturbation.
  - ✓ Predicting best time for wind farm energy utilization.
  - ✓ Integrating weather conditions for predicting various time periods like per day, per week, per month, and annual reports for wind energy generation.
  - ✓ Graphical representations and reports to support various business decisions on improving wind energy generation.
  - ✓ Balancing production and utilization of the wind energy.

PROPOSED SOLUTION

- Foremost aim of this project is to track effect of weather data to energy production.
- Particularly interested in the correlation of different components that characterize weather conditions such as wind speed, pressure, and temperature.
- As per the requirement of time constraints, presented an analysis of different time intervals:
    - ➢ One-day measurements in ten-minute intervals.
    - ➢ One-week measurements in ten-minute intervals.
    - ➢ One-month measurements in ten-minute intervals.

THEORITICAL ANALYSIS: Block diagram

Starting with
- Identify the Problem
- Analysis of Present Situation
- Importing Dataset

Data Preprocessing
- Analyze the Data
- Visualize Insights
- Correlated variable
- Important Parameter

Model Creation
- Finding Suitable Model
- Predicting Output
- Visualize Output
- Checking Accuracies
- Evaluation of model

Prediction
- Takes data
- Predict Output
- Tell best time to utilize the energy
- Give Grid result

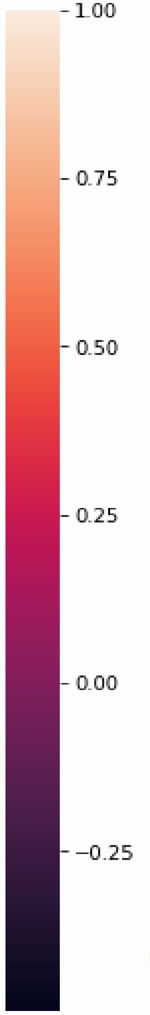
Development of Model
- Creation of Web base framework
- Wich take data and give required output

THEORITICAL ANALYSIS: Hardware / Software designing

- Type of application: IBM Watson Studio / Jupyter notebook (Anaconda) and Node-RED
- Operation system: Web frameworks IBM Cloud Watson Studio / Windows 10
- IDE: Jupyter Notebook
- Programing Language: Python 3.6
- Dataset: https://www.kaggle.com/berkerisen/wind-turbine-scada-dataset
- Libraries Version: numpy 1.15.4, pandas 0.24.1, matplotlib 3.0.2, seaborn 0.9.0, Scikit-learn 0.19.0

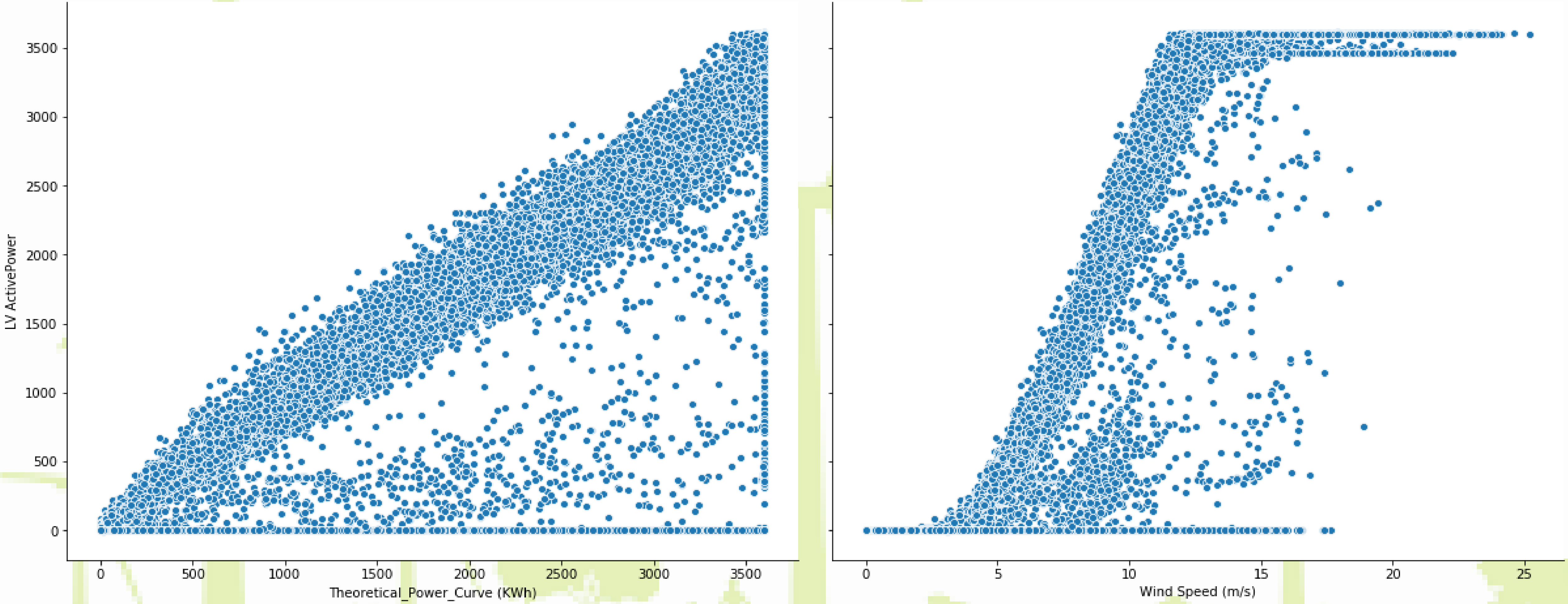
EXPERIMENTAL INVESTIGATIONS: Correlation Matrix

	Minute	Hour	Day	Month	Year	Wind Speed (m/s)	Theoretical_Power_Curve (KWh)	Wind Direction	maxtempC	windGustKmph	DewPointC	windChillC	humidity	pressure	LV ActivePower
Minute	1	6.2e-05	-0.00033	-0.00011		0.00017	-2e-05	-0.0018	-6.5e-05	-8e-05	-9.6e-05	-4e-05	-9.4e-05	4.9e-06	-2.8e-05
Hour	6.2e-05	1	-0.0019	0.0015		0.025	0.035	-0.015	0.00076	-0.0037	0.0012	0.0014	-0.0006	0.0011	0.036
Day	-0.00033	-0.0019	1	0.05		-5.8e-05	0.017	-0.0083	-0.037	0.042	-0.025	-0.041	0.032	-0.0059	-0.0075
Month	-0.00011	0.0015	0.05	1		-0.013	0.0045	-0.18	0.15	-0.032	0.21	0.18	0.037	0.26	0.031
Year															
Wind Speed (m/s)	0.00017	0.025	-5.8e-05	-0.013		1	0.94	-0.077	-0.11	0.64	-0.17	-0.16	-0.037	0.024	0.91
Theoretical_Power_Curve (KWh)	-2e-05	0.035	0.017	0.0045		0.94	1	-0.099	-0.096	0.62	-0.15	-0.15	-0.023	0.054	0.95
Wind Direction	-0.0018	-0.015	-0.0083	-0.18		-0.077	-0.099	1	-0.16	-0.16	-0.23	-0.14	-0.15	-0.21	-0.063
maxtempC	-6.5e-05	0.00076	-0.037	0.15		-0.11	-0.096	-0.16	1	-0.19	0.92	0.99	-0.49	-0.4	-0.045
windGustKmph	-8e-05	-0.0037	0.042	-0.032		0.64	0.62	-0.16	-0.19	1	-0.21	-0.28	0.12	0.024	0.57
DewPointC	-9.6e-05	0.0012	-0.025	0.21		-0.17	-0.15	-0.23	0.92	-0.21	1	0.93	-0.13	-0.41	-0.11
windChillC	-4e-05	0.0014	-0.041	0.18		-0.16	-0.15	-0.14	0.99	-0.28	0.93	1	-0.47	-0.38	-0.088
humidity	-9.4e-05	-0.0006	0.032	0.037		-0.037	-0.023	-0.15	-0.49	0.12	-0.13	-0.47	1	0.055	-0.053
pressure	4.9e-06	0.0011	-0.0059	0.26		0.024	0.054	-0.21	-0.4	0.024	-0.41	-0.38	0.055	1	0.02
LV ActivePower	-2.8e-05	0.036	-0.0075	0.031		0.91	0.95	-0.063	-0.045	0.57	-0.11	-0.088	-0.053	0.02	1

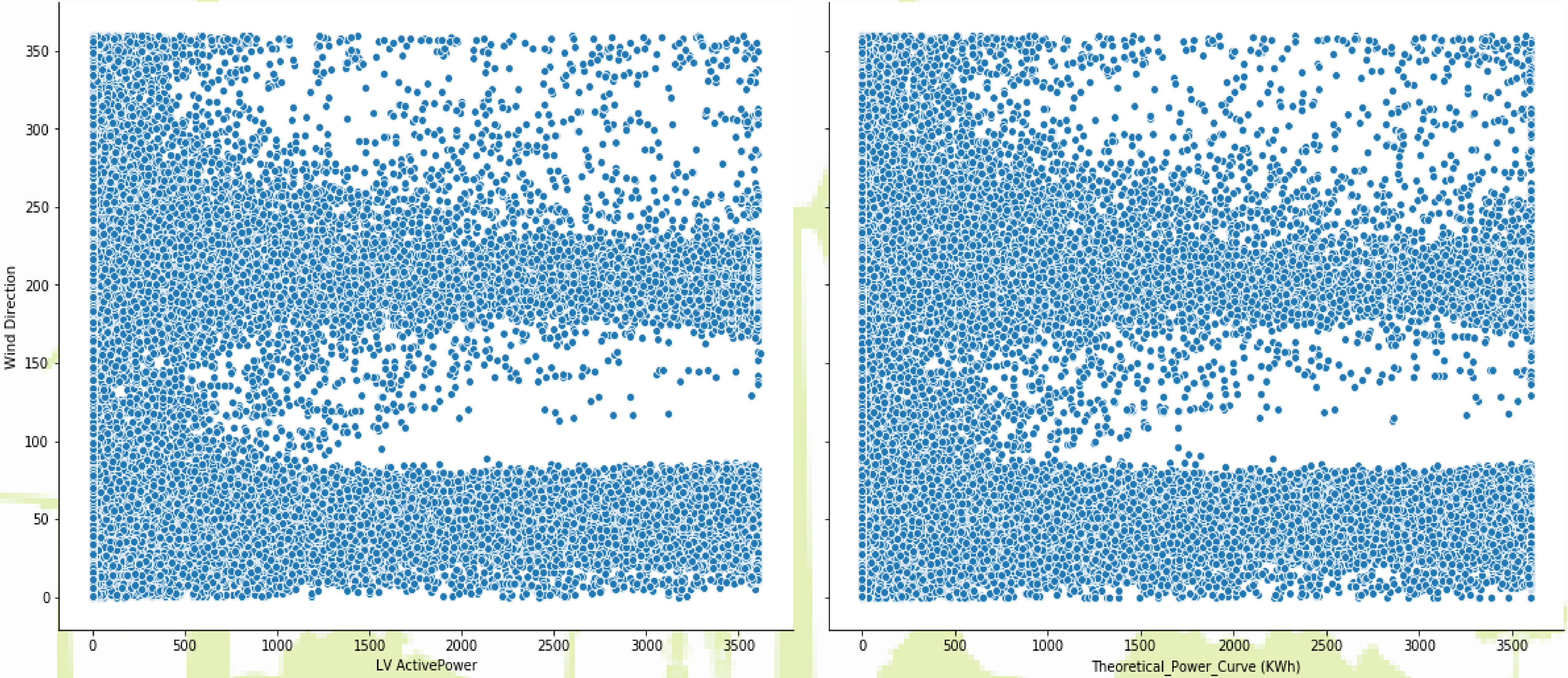


- The variable windChillC is highly correlated with maxtempC and DewPointC.
- Hence the latter once are removed as they have similar impact which can be achieved by windChillC variable.

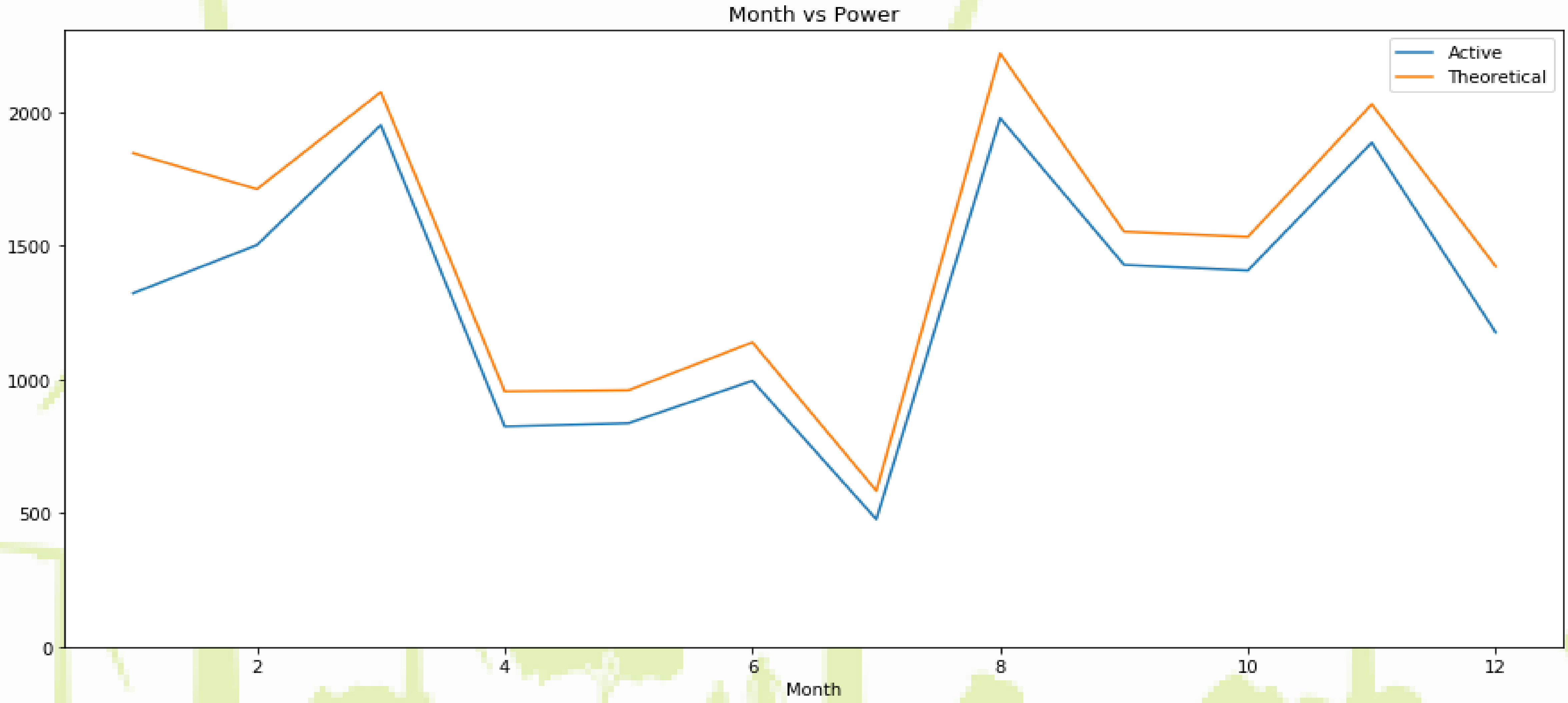
EXPERIMENTAL INVESTIGATIONS: Scatter plot of LA ActivePower over wind direction and Theoretical_Power_Curve



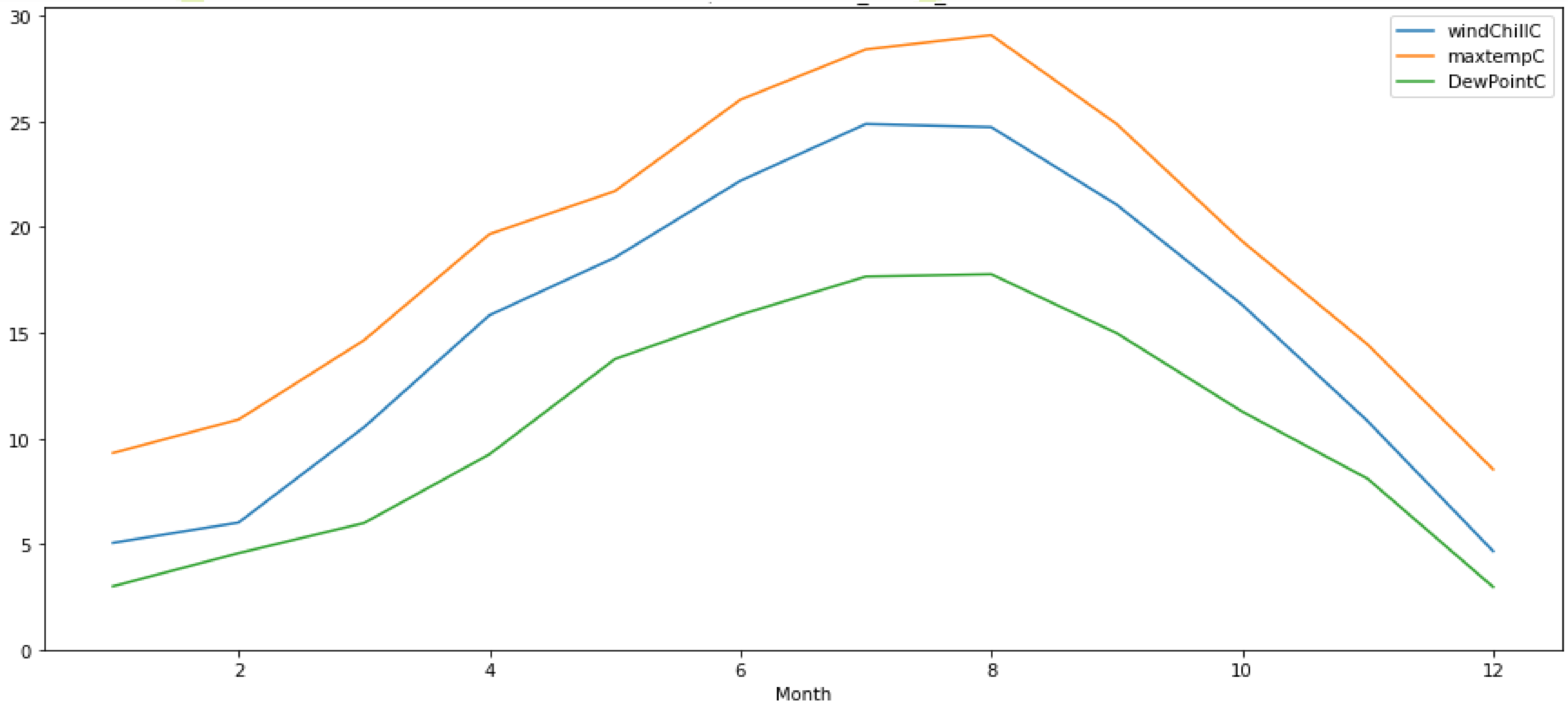
EXPERIMENTAL INVESTIGATIONS: Scatter plot of wind direction over LA ActivePower and Theoretical_Power_Curve



EXPERIMENTAL INVESTIGATIONS: Plot of LA ActivePower and Theoretical_Power_Curve over months



EXPERIMENTAL INVESTIGATIONS: Comparative plot for windChillC, maxtempC, and DewPointC over one month



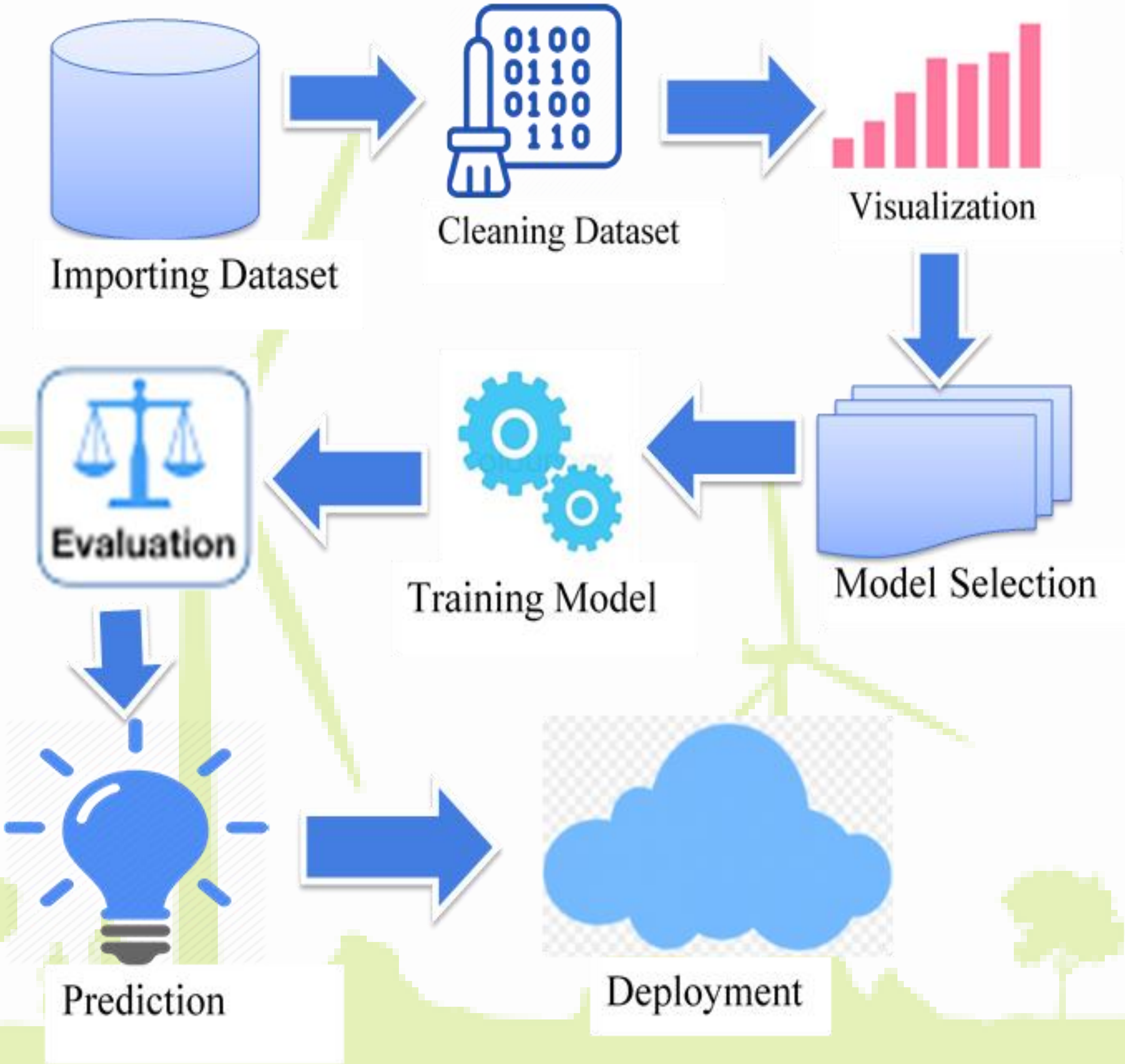
EXPERIMENTAL INVESTIGATIONS: Comparative performance of various regress models

Type of regression model	Support Vector Regression (SVR)	Linear Regression	Decision Tree Regression	Random Forest Regression	XGBoost Regression
Mean Absolute Error (MAE)	754.49	278.54	263.20	244.49	244.56
Mean Squared Error (MSE)	762843.93	313295.45	355106.95	277457.31	255542.21
Root Mean Squared Error (RMSE)	873.40	559.72	595.90	526.74	505.511
R2 Score	0.5959	0.8340	0.8119	0.8530	0.8646

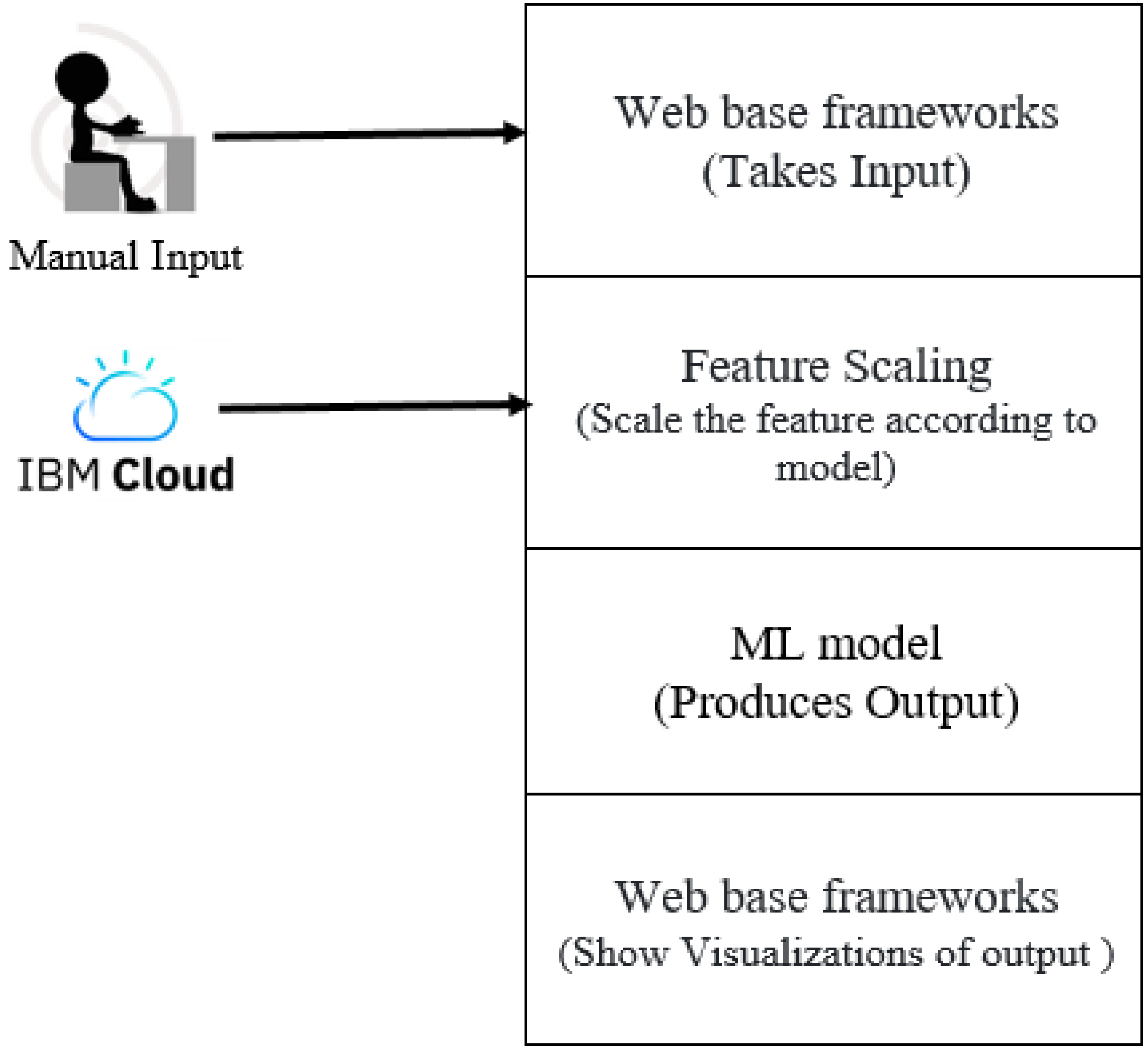
- It is noticed that XGBoost model has very low MAE, MSE and gives best mean accuracy of 0.932. Thus it is selected best suitable model for implementation on given task.
- For other performance parameters like RMSE and R2 score, XGBoots beats the other models showing very good performance on almost all the parameters.

FLOWCHART: Project modules and working

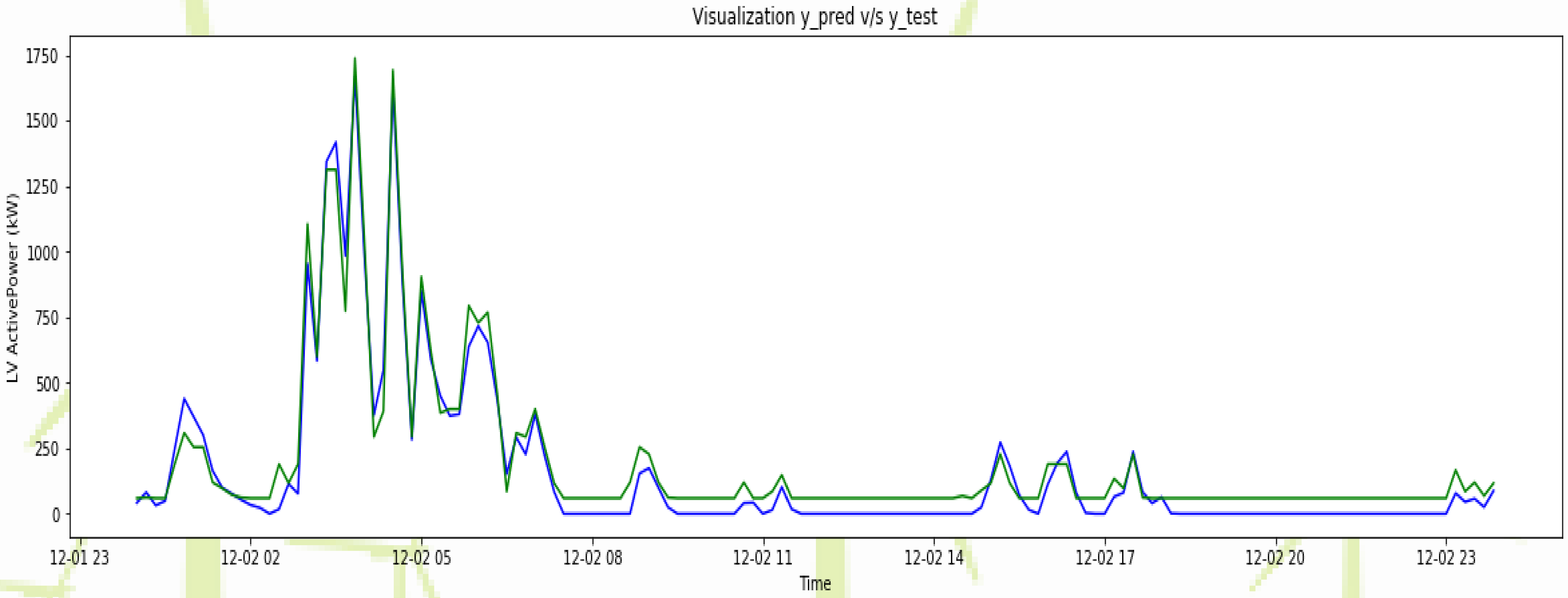
- Shows the flow of project modules implementation.
- Represents generalized project steps carried and their synchronization.



FLOWCHART: Project Web Application

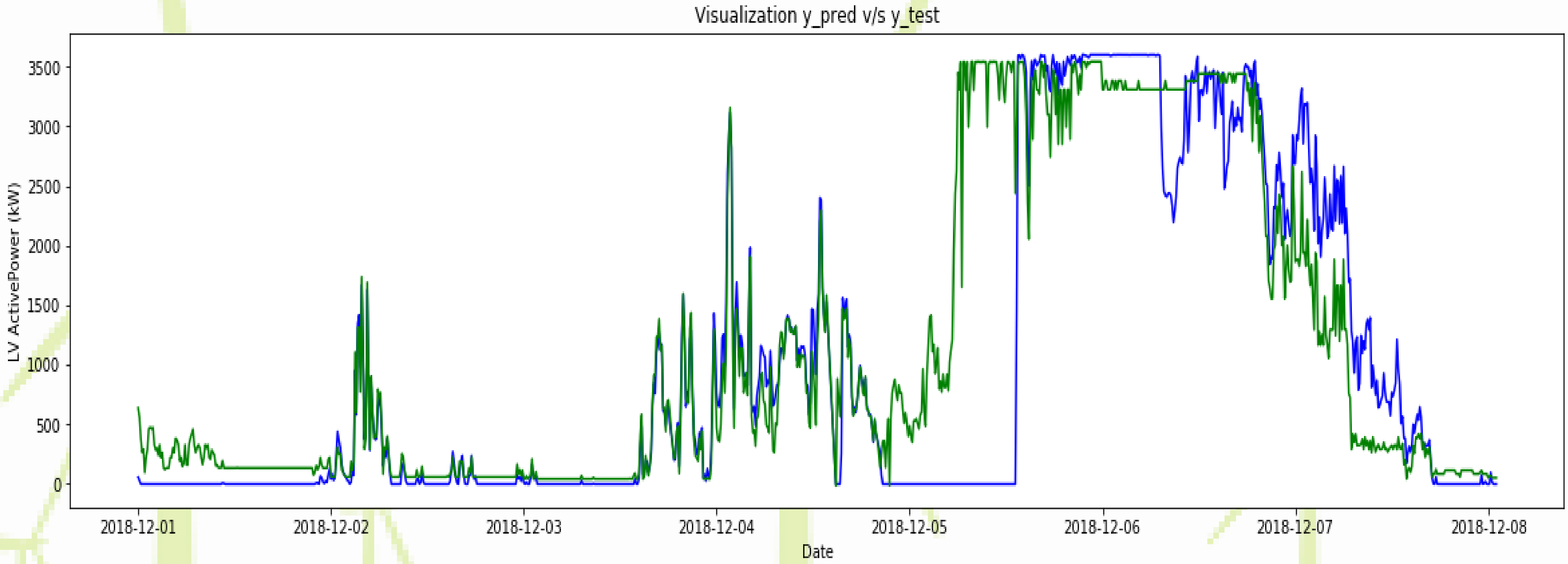


RESULTS: One-day measurements in ten-minute intervals

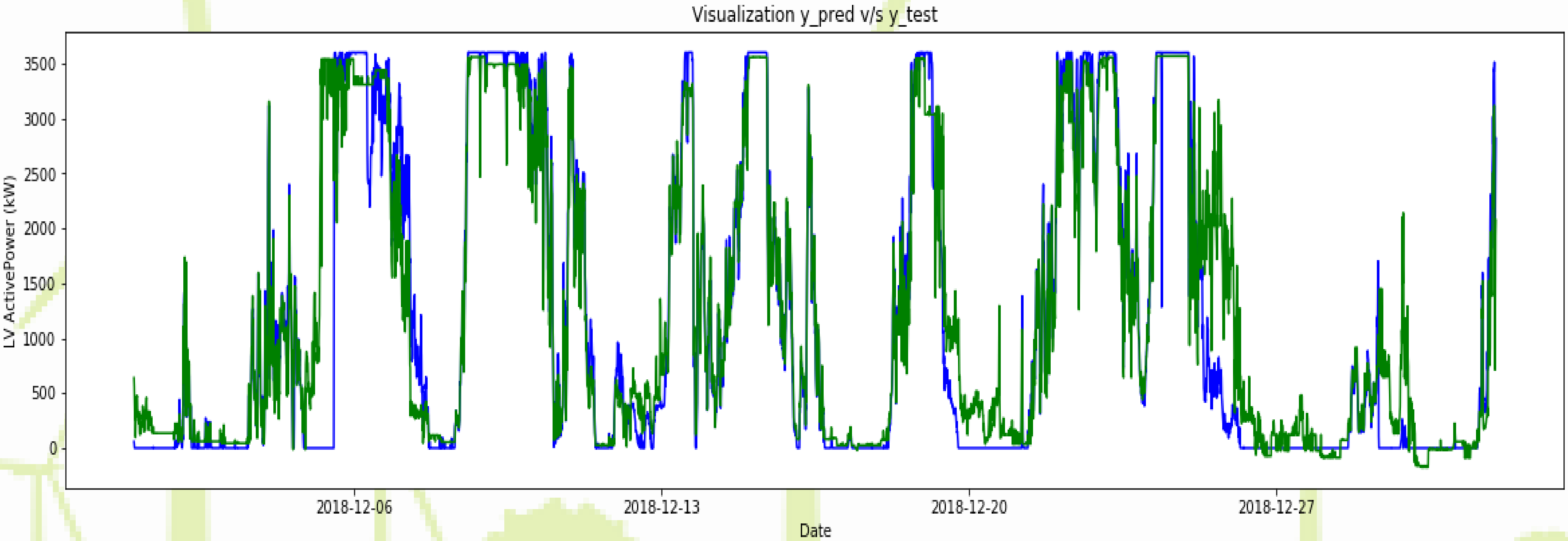


- From above result it is observed that power generation is high between 3am to 6am.
- Power generation reaches its maximum pick during said period in a day.

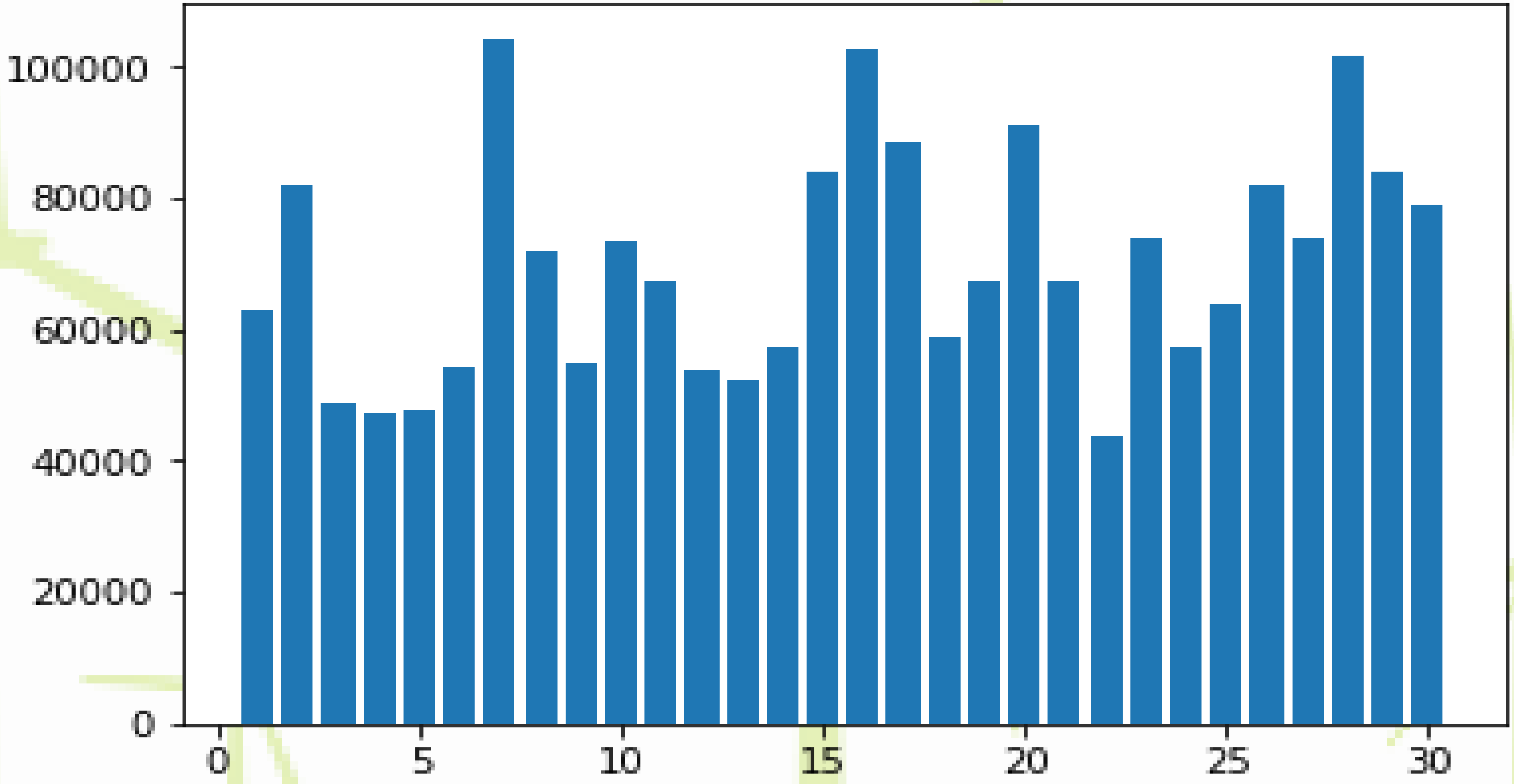
RESULTS: One-week measurements in ten-minute intervals



RESULTS: One-month measurements in ten-minute intervals

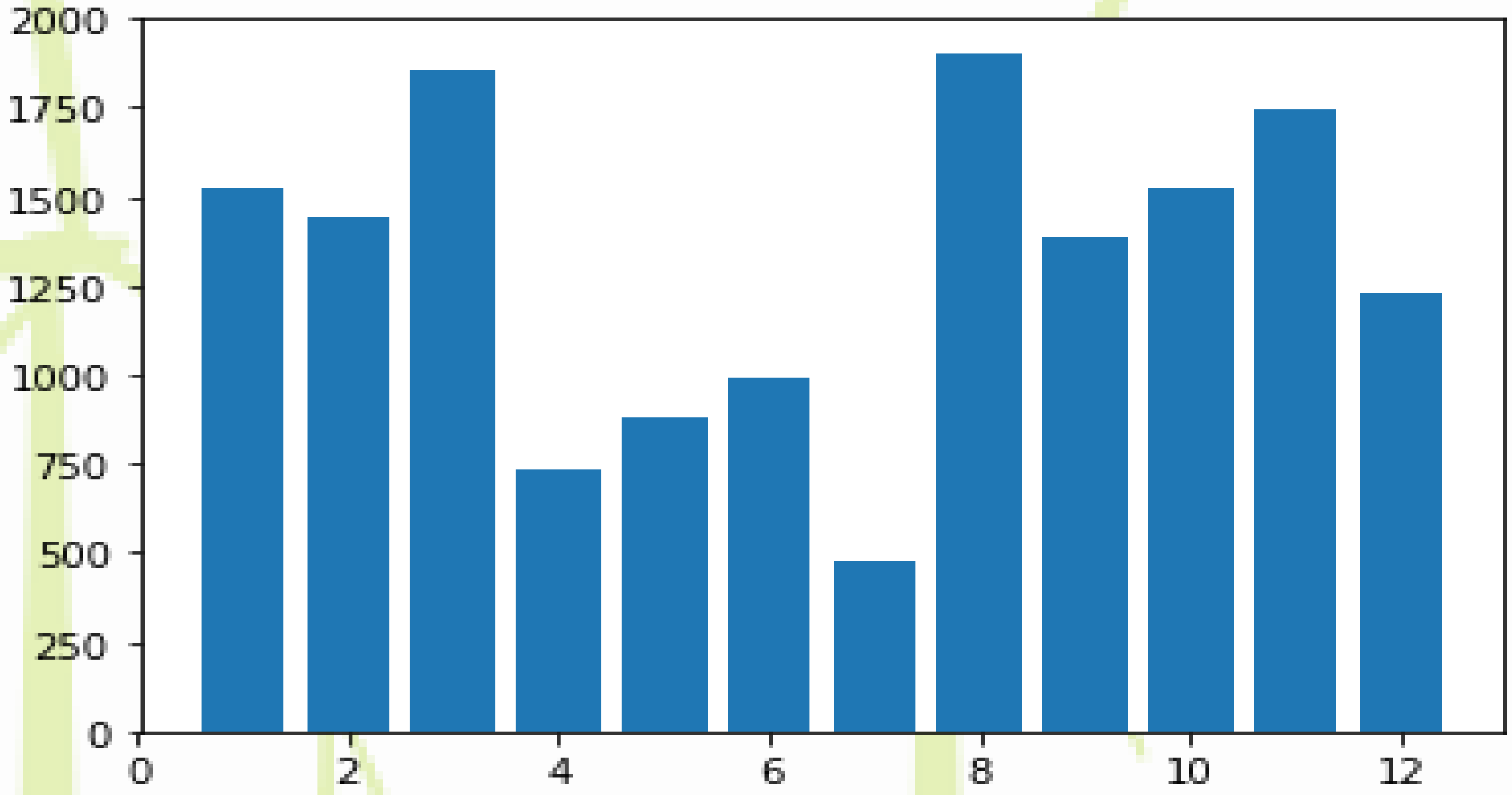


RESULTS: Day wise power generation



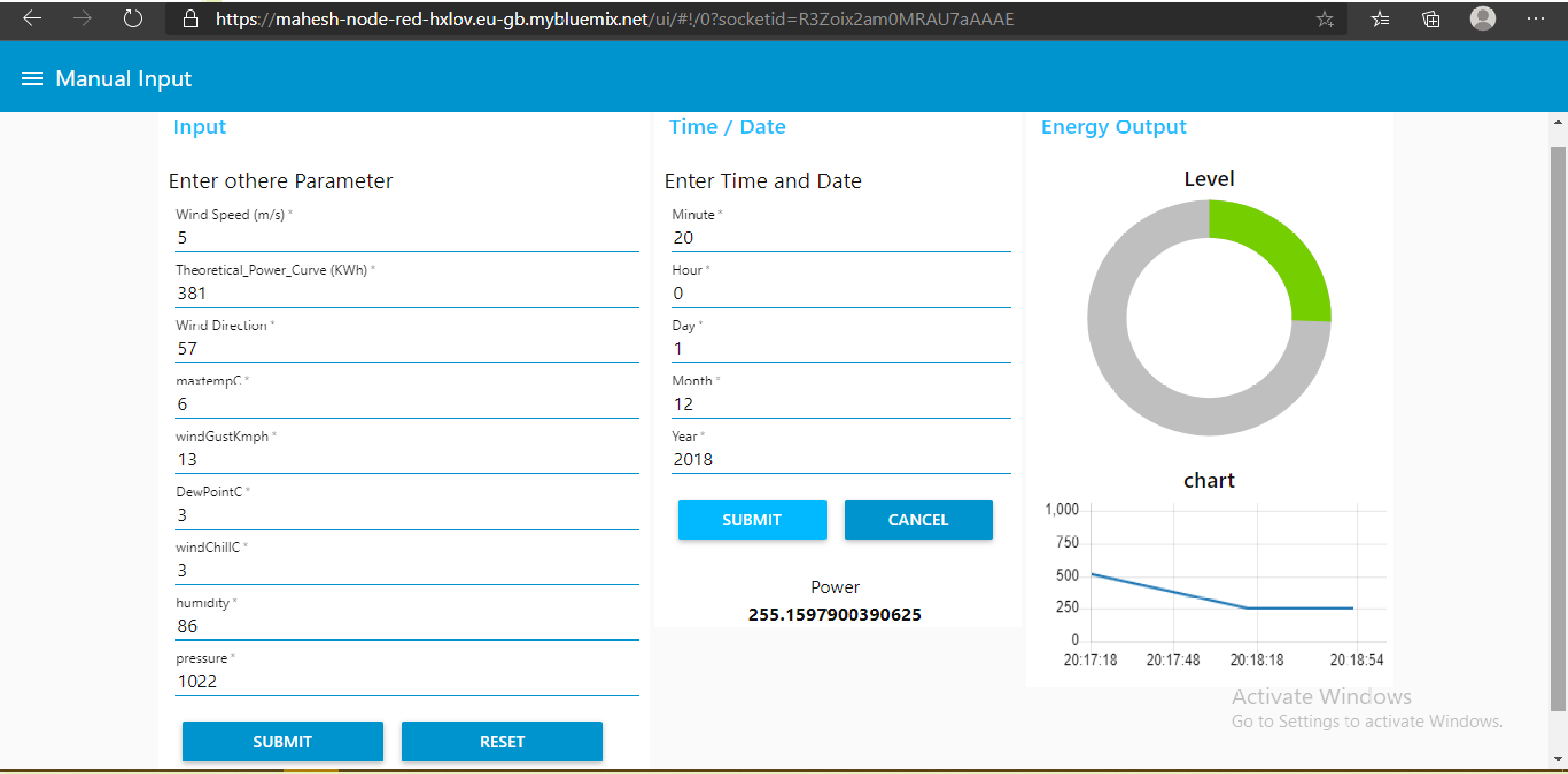
- From above figure it is noted that the power generation is highest on day-7, day-16, and day-28..

RESULTS: Month wise power generation



- The best month to utilize the energy is 8th, as the power generation is highest in the month.
- The average production of Energy = 1904.049918401606 each Minute in the 8th month.

Node-RED Web Framework



CONCLUSION

- As seen from results and discussion, the proposed algorithms give satisfying results for ten-minute measurements.
- For short term prediction, large number of training data is not required.
- Predictions are quite satisfying if model is trained on just one day if one day ahead is being predicted.
- For long term predictions, larger dataset would have to be used for training which would include data for all four seasons.

FUTURE SCOPE

- Some of the future research tasks can be targeted as:
  - ➢ To identify more environment parameters for testing their impact on wind energy generation.
  - ➢ To avail on-demand supply of wind energy.
  - ➢ To predict customer usage pattern and try to map with the wind energy generation for better business production.

BIBILOGRAPHY

[1]. Board, J., & Dashboard, C. P. S. A not-for-profit organization, IEEE is the world's largest technical professional organization dedicated to advancing technology for the benefit of humanity.

[2]. Ekaterina Vladislavleva, Tobias Friedrich, Frank Neumann, Markus Wagner, Predicting the energy output of wind farms based on weather data: Important variables and their correlation, Renewable Energy, Volume 50, 2013, Pages 236-243, ISSN 0960-1481.

[3]. Chen, N., Qian, Z., Nabney, I. T., & Meng, X. (2013). Wind power forecasts using Gaussian processes and numerical weather prediction. IEEE Transactions on Power Systems, 29(2), 656-665.

[4]. [Online] https://www.energy.gov/eere/wind/2018-wind-market-reports.

[5]. [Online] Dataset-Historical Weather API and Kaggle