TEST Engineering & Management

January - February 2020
ISSN: 0193 - 4120 Page No. 6321 - 6326

Regression Model Accuracy Measurement and Evolution for Sample Data for Hybrid Solar and Wind Power

Assitant Prof. Dr. Ravirajsinh S. Vaghela[1,a]
[1]*R. B. Institute of Management and Computer Studies, Gujarat TechnologcialUniversity*

Assitant Prof. Dr. Siddharth Joshi [2,b]
[2]*P. D. P. U, Gandhinagar*

Article Info
Volume 82
Page Number: 6321 - 6326
Publication Issue:
January-February 2020

Article History
***Article Received**: 18 May 2019*
***Revised**: 14 July 2019*
***Accepted**: 22 December 2019*
***Publication**: 30 January 2020*

Abstract:

This paper purely based on the machine learning regression model accuracy checking for the sample dataset of the power generation from wind and solar energy conversion system in standalone mode. The work is also focused on the combined power generation from wind and solar with incorporation of battery bank. For that the data is generated for the hybrid renewable energy system with wind, solar PV and battery; through simulation software. The wind speed for wind turbine, radiation and temperature for solar PV is taken in to the consideration for Ahmedabad, Gujarat, India. Through machine learning training the accuracyhas checked for the metrics to provide the proper insight for the dataset of the hybrid renewable energy power generation system and its prediction forgiven dataset of monthly basis.

Keywords:*Hybrid renewable energy system, Machine learning, Solar PV, Wind*

I. Introduction

The growth rate of renewable energy generation is fast because of the global warming issues and exhaustible nature of the fossil fuel and conventional energy of sources as well. The energy demand is rising day by day. For that people are focusing on renewable energy sources. The wind and solar energy systems are the powerful market players to overcome these demand in most significant way.

The data is generated for the machine learning regression model [1, 2, 3] through PSIM ® 11.1.3 software. The system used in for testing is depicted in fig. 1. The system comprises of 3kW of PV array and 3.5kW of wind energy conversion system. For the reliability point of view (as wind and solar resources are stochastic in nature) the battery energy storage system of 72V, 25Ah for back up. The closed loop simulation with gain scheduling proportional integral controller is used to tune the PI controller and to avoid the disadvantages of fixed gain PI controllers. The data used as an input to perform the simulation are wind speed, solar radiation, temperature of solar PV and initial state of charge of battery bank. The wind speed, solar radiation and temperature for PV is used for monthly variations for the Ahmedabad, Gujarat, India. The input data is available in the literature.

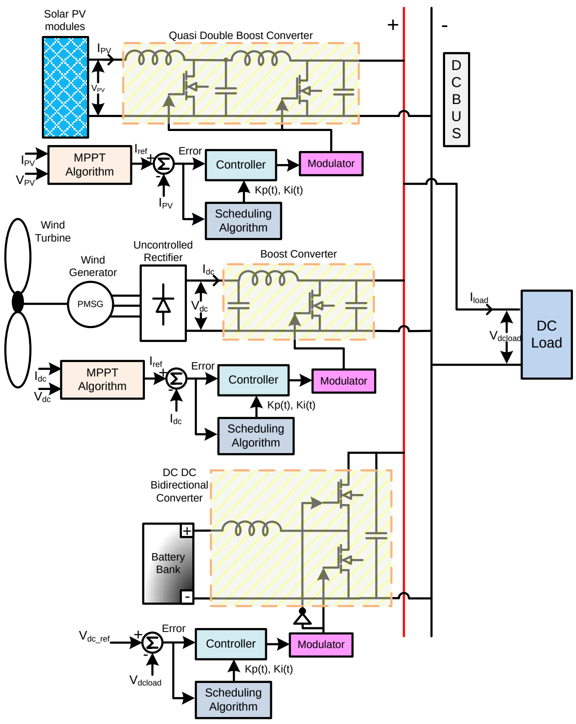


Fig. 1 Simulation test system for data generation

II. Review of Literature

In [4], author advocated that Regression analysis mainly focused on develop such regression model that can give predicated result on the basis of given dependent input feature and output for the training. They proposed the new model of patent regression.

In [5], author advocated that if sample sized is small or medium of any dataset, for the regression the true method is AICC, gives good model choice when the dimension is finite. In [6], author advocated model comparison and regression analysis two different paradigm. Though both the process applied on the same dataset. They note that for the linear regression model selection can be done on small sample. In [7], author suggested that time series multivariate analysis, model score comparison for the best result indeed. They analyzed and conclude the best linear regression model is LeastMedSq over SMOreg function. In [8], author suggested that the MGM (1, m) with optimized background values useful to remove the random errors of the data and the combined prediction model together with the multiple linear regression model has good effects for prediction. In [2], author referees that K-nearest neighbor (k-NN) is modest and old-style nonparametric method (Bishop, 1995; Manocha&Girolami, 2007). In [9], author suggested that kNN is an instance-based learner and Knn successfully work in StatLog Project. In [1], author proved by experimental study KNN algorithm outperformed on different datasets. Author also experiment KNN by Class Based Weighted KNN. In [3], author suggested that Knn algorithm internal mechanism calculate distance of training data on that basis predicate the output. In [10], author advocated that performance of SVR highly depends on its parameters. And author study existing SVR algorithm and proposed firefly algorithm based memetic algorithm to judge best depend parameter which give more accurate result than other existing SVR algorithm. SVR model can yield more accurate prediction result proved by experimental result. In [11], SVM provide more accurate result though its complex in nature. Overfitting problem can be overcome by SVM model. SVM also give result if data is not linear in nature by kernel concept. In [3], SVM calculate a set of weight of features based on a transformation of the feature space. SVM internal main process is dividing and deciding best hyperplane.

III. Methodology

For experimenting and testing we choose python, jyupter and sklearn library to load, feature selection for train and testing the model. Step 1 to aggregate all the month sample data of 15 seconds. After making single .csv file of 12 month data. Load .csv file into python using panda's library and segregate dependent variable of solar and wind in different data frame. Following are the figure populated after segregate wind and solar data.

The .csv file is generated after getting by running the simulation shown in fig. 1 in PSIM software. The simulation analysis is done using real time data for the Ahmedabad, Gujarat, India location available in [12, 13, 14, 15]. The data of wind speed, radiation and temperature of PV used as the input parameters of the simulation. The input parameters are depicted in appendix 1 at the last.

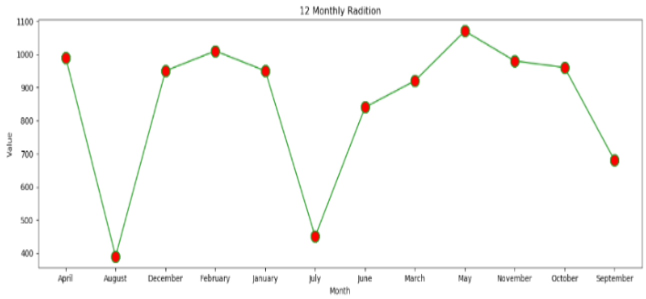


Figure 2: 12 month wise sample solar radiation as input

From Figure 2 we can rectify aggregate solar of the 12 month which will main feature for the predication of solar power.

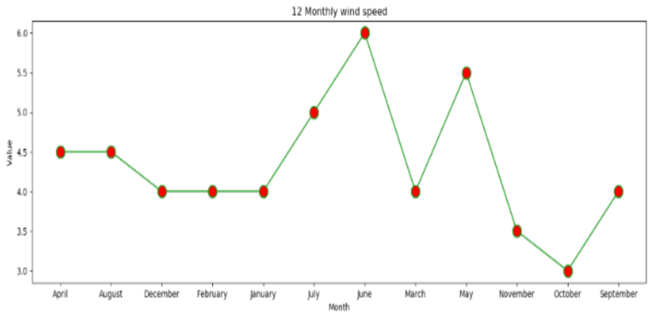


Figure 3: 12 month wise sample wind speed as input

From Figure 3 we can rectify aggregate wind power of the 12 month which will main feature for the wind power generation.

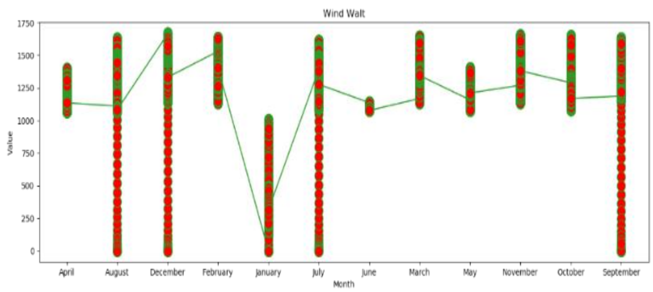


Figure 4: 12 month wise sample of solar Walt

From Figure 4 label which is indicate sample wind watt of 12 month wind speed output.

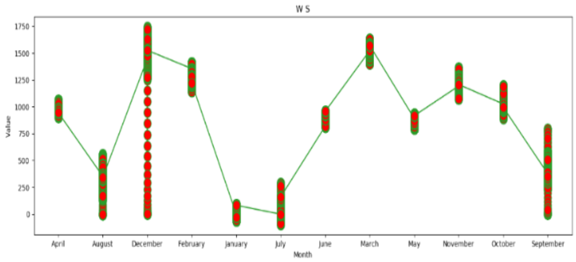


Figure 5: Wind power label 12 month sample

From Figure 5 label which is indicate sample watt of 12 month.

After visualizing data and resultant class next step is to separating data between trains and testing via train_test_split methodology in python. By which our model train on 80 percent of data and 20 percent data is hidden from our train model to test its accuracy. Here same data trained with Liner regression algorithm, SVM regression algorithm, and Random forest. After testing comparison tables of important model metrics drawn [16, 17, and 18].

IV. Result and Analysis

1) Separately trained for solar prediction based on features

Algorithm / Evolution of model	mean squared error	mean absolute error	Coefficient correlation
Linear Regression	110524.89	231.59	0.568
SVM Regression	81727.67	206.87	-0.16
Random Forest Regression	4578.878	47.18	0.982
KNN Regression	5193.32	28847.80	121.23

Table 1: Solar features and class only for model evolution Metrics using sklearn library

From the Table 1 : we can see the result of Solar trained and tested model in which Random forest mean absolute error & mean squared error is less compared to SVM, Linear Regression, KNN regression model so we can say for the solar power prediction model of Random forest is better prediction model.

2) Separately trained for Wind power prediction based on features

Algorithm / Evolution of model	mean squared error	mean absolute error	Coefficient Correlation
Linear Regression	66693.49	157.33	0.050
SVM Regression	77308574072.5	267754.82	-299559.49
Random Forest Regression	19231.338	97.78	0.72
KNN Regression	28847.80	121.23	0.58

Table 2: Wind features and class only for model evolution Metrics using sklearn library.

From the Table 2 : we can see the result of Wind trained tested model in which Random forest mean absolute error & mean squared error is less compared to SVM, Linear Regression, KNN regression model so we can say for the solar power prediction model of Random forest is better prediction model[16, 17, 18].

So we did separate work experiment on dataset wind and solar features. We could find better result from Random forest which having wining strike over other supervised machine learning algorithm.

After this we did experiment on combined features of wind and solar radiation and trained our dataset model with same algorithms in Weka Tools.

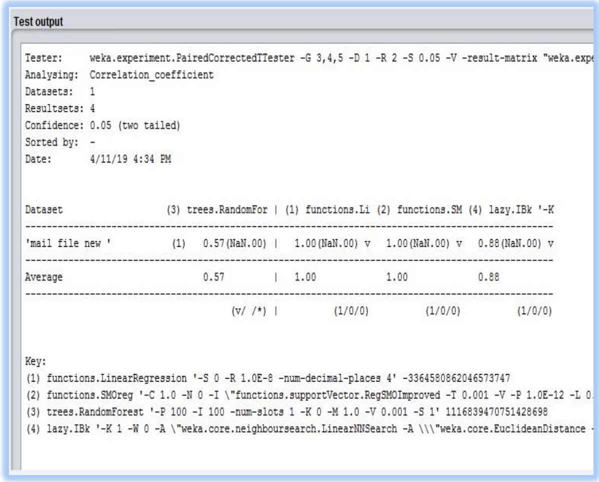


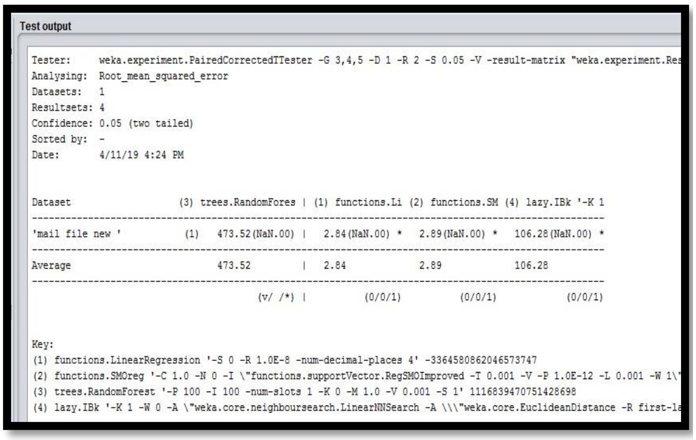
Figure 6: Average of Correlation_ coeffient of trained model with Linear Regression, SMoreg, Random Forest, and KNN

From Figure 6 average of correlation of Random Forest have a little “v”. This means that the difference in the correlation coefficient for these algorithms compared to linear regression, Support vector and KNN-IBK is more statistically significant [19, 20, 21, 22,23].

Algorithm / Evolution of model	Average Score of Correlation Coefficient
Linear Regression	1
SVM Regression	1
Random Forest Regression	0.57
KNN Regression	0.88

Table 3 Model Evolution Correlation Coefficient comparison metrics of Combine features

From Table we can see average score of correlation coefficient the Random Forest is less which is more relevant indicator to better than Linear Regression, SVM Regression, KNN IBK Model.



```
Test output

Tester:     weka.experiment.PairedCorrectedTTester -G 3,4,5 -D 1 -R 2 -S 0.05 -V -result-matrix "weka.experiment.Res
Analysing:  Root_mean_squared_error
Datasets:   1
Resultsets: 4
Confidence: 0.05 (two tailed)
Sorted by:  -
Date:       4/11/19 4:24 PM


Dataset                  (3) trees.RandomFores | (1) functions.Li (2) functions.SM (4) lazy.IBk '-K 1
---------------------------------------------------------------------------------------------------
'mail file new '          (1)   473.52(NaN.00) |   2.84(NaN.00) *   2.89(NaN.00) *   106.28(NaN.00) *
---------------------------------------------------------------------------------------------------
Average                         473.52         |   2.84             2.89             106.28
---------------------------------------------------------------------------------------------------
                                       (v/ /*) |          (0/0/1)          (0/0/1)          (0/0/1)


Key:
(1) functions.LinearRegression '-S 0 -R 1.0E-8 -num-decimal-places 4' -3364580862046573747
(2) functions.SMOreg '-C 1.0 -N 0 -I \"functions.supportVector.RegSMOImproved -T 0.001 -V -P 1.0E-12 -L 0.001 -W 1\"
(3) trees.RandomForest '-P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1' 1116839470751428698
(4) lazy.IBk '-K 1 -W 0 -A \"weka.core.neighboursearch.LinearNNSearch -A \\\"weka.core.EuclideanDistance -R first-la
```

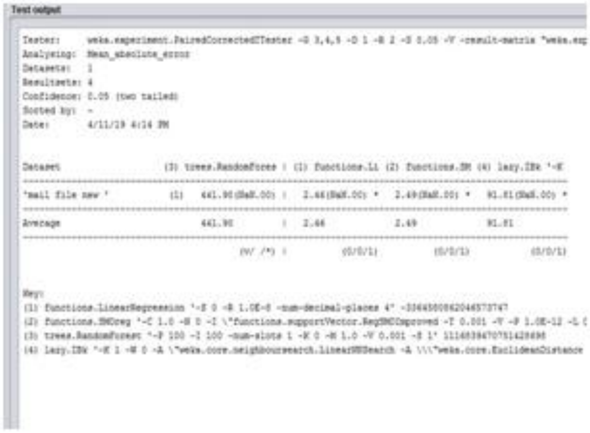
Figure 7: Weka experiment's result Average Root mean squared error of Different Regression train model

From Figure 7 average of correlation of Random Forest have a little "v". This means that the difference in the Root mean square error for these algorithms compared to linear regression, Support vector and KNN-IBK is more statistically significant **[**19, 20, 21, 22,23].

Algorithm / Evolution of model	**Avg. Root mean square error**
Linear Regression	2.84
SVM Regression	2.89
Random Forest Regression	475.52
KNN Regression	106.28

Table 4: Model Evolution Root mean square error comparison metrics of Combine features

From Table 4 we can see average root mean square error score of the Random Forest is more significant which is more relevant indicator to better than Linear Regression, SVM Regression, KNN IBK Model.



```
Test output

Tester:     weka.experiment.PairedCorrectedTTester -G 3,4,5 -D 1 -R 2 -S 0.05 -V -result-matrix "weka.exp
Analysing:  Mean_absolute_error
Datasets:   1
Resultsets: 4
Confidence: 0.05 (two tailed)
Sorted by:  -
Date:       4/11/19 4:14 PM


Dataset                  (3) trees.RandomFores | (1) functions.Li (2) functions.SM (4) lazy.IBk '-K
---------------------------------------------------------------------------------------------------
'mail file new '          (1)   441.90(NaN.00) |   2.46(NaN.00) *   2.49(NaN.00) *   91.81(NaN.00) *
---------------------------------------------------------------------------------------------------
Average                         441.90         |   2.46             2.49             91.81
---------------------------------------------------------------------------------------------------
                                       (v/ /*) |          (0/0/1)          (0/0/1)          (0/0/1)


Key:
(1) functions.LinearRegression '-S 0 -R 1.0E-8 -num-decimal-places 4' -3364580862046573747
(2) functions.SMOreg '-C 1.0 -N 0 -I \"functions.supportVector.RegSMOImproved -T 0.001 -V -P 1.0E-12 -L (
(3) trees.RandomForest '-P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1' 1116839470751428698
(4) lazy.IBk '-K 1 -W 0 -A \"weka.core.neighboursearch.LinearNNSearch -A \\\"weka.core.EuclideanDistance
```

Figure 8: Mean absolute error Metrics for Regression Model

From Figure 8 average of correlation of Random Forest have a little "v". This means that the difference in the mean absolute error for these algorithms compared to linear regression, Support vector and KNN-IBK is more statistically significant [19, 20, 21, 22,23].

Algorithm / Evolution of model	**Avg. Mean absolute error**
Linear Regression	2.46
SVM Regression	2.49
Random Forest Regression	91.01
KNN Regression	2.49

Table 5: Mean Evolution Root mean square error comparison metrics of Combine features

From Table 5 we can see average mean absolute error score of the Random Forest is more significant which is more relevant indicator to better than Linear Regression, SVM Regression, KNN IBK Model.

V. CONCLUSION

The generated data from the simulation tool is used in the machine learning process. The data is generated from the test system for the training and testing on the dataset with various machine learning algorithm is done. The test is performed with for KNN regression, Linear Regression, SVM Regression, and Random forest Regression. Among that Radom forest is best suited for prediction.

January - February 2020
ISSN: 0193 - 4120 Page No. 6321 - 6326

REFERENCES

[1]. Tsai, C. F., Hsu, Y. F., Lin, C. Y., & Lin, W. Y. (2009). Intrusion detection by machine learning: A review Expert Systems with Applications.

[2]. Wagh, Sumit& V Walke, P. (2017). REVIEW ON WIND-SOLAR HYBRID POWER SYSTEM.Philips D (2000).

[3]. Al Badwawi, Rashid &Abusara, Mohammad &Mallick, Tapas. (2015). A Review of Hybrid Solar PV and Wind Energy System. Smart science. 3. 127-138. 10.6493/SmartSci.2015.324.

[4]. El Bassam, N., &Maegaard, P. (2004). Integrated renewable energy for rural communities: Planning guidelines, technologies and applications. Elsevier.

[5]. Hurvich, C. M., & Tsai, C. L. (1989). Regression and time series model selection in small samples. Biometrika, 76(2), 297-307.

[6]. Hurvich, C. M., & Tsai, C. L. (1990). The impact of model selection on inference in linear regression. The American Statistician, 44(3), 214-217.

[7]. Kavitha, S., Varuna, S., &Ramya, R. (2016, November). A comparative analysis on linear regression and support vector regression. In 2016 Online International Conference on Green Engineering and Technologies (IC-GET) (pp. 1-5). IEEE.

[8].Xiong, P., Dang, Y., Wu, X., & Li, X. (2011). Combined model based on optimized multi-variable grey model and multiple linear regression. Journal of Systems Engineering and Electronics, 22(4), 615-620.

[9]. Bichler, M., & Kiss, C. (2004). A comparison of logistic regression, k-nearest neighbor, and decision tree induction for campaign management. AMCIS 2004 Proceedings, 230.

[10].Hu, Z., Bao, Y., &Xiong, T. (2013). Electricity load forecasting using support vector regression with memetic algorithms. The Scientific World Journal, 2013.

[11].Dubey, H. (2013). Efficient and accurate kNN based classification and regression. A Master Thesis Presented to the Center for Data Engineering, International Institute of Information Technology, Hyderbad-500, 32.

[12].Joshi, S., Pandya, V., &Sant, A. V. (2018). Gain Scheduling Algorithm-Based Control of Renewable Energy Systems for Hybrid Standalone DC Grid. Iranian Journal of Science and Technology, Transactions of Electrical Engineering, 42(3), 327-342.

[13].https://data.gov.in/catalog/progress-report-village-electrification (2016), progress report of village electricfication, accessed on September, 2016.

[14]. World Energy Outlook (2016) – Electricity AccessDatabase,www.worldenergyoutlook.org, accesed on December 2016.

[15]. Energy Statistics Central Statistics Office, Government of India, report by GOI, www.mospi.gov.in, from 2012 to 2016 yearly report.

[16]. Kramer, O. (2016). Scikit-learn. In Machine learning for evolution strategies (pp. 45-53). Springer, Cham.

[17].Hamid, N. A., &Sjarif, N. N. A. (2017). Handwritten recognition using SVM, KNN and neural network. arXiv preprint arXiv:1702.00723.

[18].Garreta, Raúl, and Guillermo Moncecchi. Learning scikit-learn: machine learning in python. Packt Publishing Ltd, 2013.

[19]. Kaur, A., & Chopra, D. (2016). Reasons for Non-Applicability of Software Entropy Metrics for Bug Prediction in Android. World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering, 10(6), 1170-1175.

[20]. García-Gutiérrez, J., Martínez-Álvarez, F., Troncoso, A., &Riquelme, J. C. (2015). A comparison of machine learning regression techniques for LiDAR-derived estimation of forest variables. Neurocomputing, 167, 24-31.

[21]. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. ACM SIGKDD explorations newsletter, 11(1), 10-18.

[22].Treiber, N. A., Heinermann, J., & Kramer, O. (2016). Wind power prediction with machine learning. In Computational Sustainability (pp. 13-29). Springer, Cham.

[23].Vaghela, R. S., Gonsai, A., &Gami, P. (2017, March). Measurement (Data Mining) of Real Mobile Signals Data in Weka Tools for Interference Detection. In International Conference on Information and Communication Technology for Intelligent Systems (pp. 654-661). Springer, Cham.