International Journal of Computer Sciences and Engineering Open Access
Survey Paper Vol. 9, Issue.9, September 2021 E-ISSN: 2347-2693

Data Mining Techniques for Estimation of Wind Speed Using Weka

B. Hari Mallikarguna Reddy[1*], S. Venkatramana Reddy[2], B. Sarojamma[3]

[1,3]Department of Statistics, S.V. University, Tirupati-517502, Andhra Pradesh, India
[2]Department of Physics, S.V. University, Tirupati-517502, Andhra Pradesh, India

**Corresponding Author: bhmreddy@gmail.com, Mobile: 9989294205*

DOI: https://doi.org/10.26438/ijcse/v9i9.4851 | Available online at: www.ijcseonline.org

Received: 18/Sept/2021, Accepted: 20/Sept/2021, Published: 30/Sept/2021

Abstract: Now a day's neural network plays a vital role in analyzing, interpreting and fitting models. In this paper by taking wind speed as dependent variable and minimum temperature, maximum temperature, visibility, temperature date and time as independent variables, we fitted. M5P, SMO Regression and zero regression models and CV parameter selection criteria is also used for above three models. For computational purpose WEKA Software is used. By measures of accuracy like mean absolute error, root mean square. Relative absolute error, root relative squared error are used to select the best model and also rank them.

Keywords: Wind speed, Zero regression, M5P, SMO regression, WEKA.

I. INTRODUCTION

Wind power or wind energy plays a vital role in Agriculture and solar energy source. Wind speed is high when temperature is high and low when temperature is low. Number of research papers are in the literature depends upon the wind speed and solar energy. Ramedani *et al.*[1] published a paper entitled "Modeling solar energy potential in a Tejran province using artificial neural networks". In this paper they used artificial neural networks (ANNs) for modeling and estimation global solar radiation by using few inputs. They used the variables such as maximum and minimum temperature, sunshine duration, day light hours, extraterrestrial radiation and number of days in the year. The main tool is one hidden layer multi layer Perceptron with 37 neurons in it. Cucumo et al.[2] written a paper that "experimental testing of models for estimation of hourly solar radiation on vertical surfaces".

Somaieh Ayalvary*et* et al.[3] describes "select the most relevant input parameters using WEKA for model forecast solar radiation based on Artificial Neural Networks". In this paper, they discussed solar radiation energy forecasts using traditional models and Artificial neural networks using solar warm Waikate Environment for Knowledge analysis (WEKA). The parameters includes latitude, longitude, maximum wind speed, average temperature, average minimum air temperature, sunshine, monthly rainfall, maximum rainfall in a day for different places of hilan. Khatib et al.[4]explains "A review of solar energy modeling" and Azeez et al.[5] gave "Artificial neural network estimation of global solar radiation using metrological parameters in Gusau, Nigeria. P. Usha Sri and B.Narasimha Swamy [6] explains about "Wireless Atmospheric Data Logger for a Sensor Network". K.V. Shende et al.[7] gives research article on "Artificial Neural Network Model for Prediction of Latent Heat Flux over Bay of Bengal".

II. METHODOLOGY

There are numerous Machine learning algorithms in the literature and some of the important and popular algorithms based on Regression are linear regression, K-nearest neighbours, decision tree, support vector machines and multi-layer perception.

A. SMO Regression

Support vector machine regression is a multi-class classification and regression problem. In support vector regression machine, input variable is numerical value and it converts automatically nominal values to numerical values. Normalization of input is being carried out before using it. Steps to carry out SMO Regression in WEKA are

Step-1: Go to weka Software
Step-2: Select data by using open file
Step-3: Go to classify and sem select SMO Reg from functions
Step-4: Click start botton to run the algorithm

B. M5P

M5P tree has mainly contains three steps, they are building the tree, pruning tree and smoothing process. Generally M5P tree algorithm enumerates any numbers of enumerated attributes and transforms into binary variables before tree construction. Basic tree of M5P is formed using the splitting criterion. Steps to Calculate M5P in WEKA are

Step-1: Open WEKA, select data from open file of preprocess.
Step-2: Select attributes you want.
Step-3: Go to classify from that "choose" to M5P from "trees".
Step-4: Click on start button to start the algorithm.

C. CV parameter selection

This met a classifier that can optimise over an arbitrary number of parameters, with only one draw back. If we want to confine the data with in limits it is very beneficial. Step by step procedures for calculation of CV parameter selection in WEKA is as follows:

Step-1: Open weka, select data from open file options.
Step-2: Select attributes from data.
Step-3: Go to classify from that "Choose" to CV Parameter selection from Meta.
Step-4: Click on to select appropriate method you want by seeing place cursor on choose opposite dialog box.
Step-5: Click to "start" algorithm.

D. Zero R

Zero R means zero regression, it is always building model with cross validation coefficient zero. This method is usually used as a reference point for comparing with other regression methods. It is the simplest classification method which relies on the target and ignores all predictors. It always predicts the majority class. Step to calculate zero R in weka is as follows:

Step-1: Open weka, select data from open file options.
Step-2: Select attributes from data.
Step-3: Go to classify from that "choose" to zero R from rules
Step-4: Click start button to start the algorithm process.

E. Correlation coefficient

Numerical measures of relation between two variables are correlation coefficient. There are 5 types of correlation coefficients according to these values.

1. . If correlation coefficient (CC) is exactly equal to 1 then it is perfect positive correlation
2. If correlation coefficient (CC) is exactly equal to -1 then it is perfect negative correlation between variables
3. if CC is >0.5, then there exists positive relation between variables.
4. if CC is <0.5, then there exists negative relation between variables.
5. if CC is '0', the two variables are independent, no correlation between them.

F. Measures of accuracy

i) MAE : It is arithmetic mean of absolute difference of prediction value and true value.

$$MAE = \frac{\sum_{i=1}^{n}(y_i - x_i)}{n} = \frac{\sum_{i=1}^{n}|e_i|}{n}$$

where y_i = Prediction value
x_i = True value
e_i = Error
n = Number of observations

III. ROOT MEAN SQUARE ERROR (RMSE) OR ROOT MEAN SQUARE DEVIATION

The following steps are used for estimation of RMSE.

Step 1: Calculate error using true value (x_i) and estimated values ($\widehat{x_i}$)
Step 2: Square each and every error values ($x_i - \widehat{x_i}$)
Step 3: Sum all squared errors i.e. $\sum_{i=1}^{n}(y_i - \widehat{x_i})^2$
Step 4: Sum divided with number of observations (n) i.e. $\sum_{i=1}^{n}\frac{(y_i-\widehat{x_i})^2}{n}$
Step 5: Take positive square root of step 4 gives RMSE i.e. $\sqrt{\frac{\sum_{i=1}^{n}(y_i-\widehat{x_i})^2}{n}}$

A. *Relative Absolute Error (RAE)***:** It is a ratio of mean error to errors produced by native model.

$$RAE = \frac{[\sum_{i=1}^{n}(u_i - x_i)^2]^{1/2}}{[\sum_{i=1}^{n} x_i{}^2]^{1/2}}$$

x_i = actual values x_i
U_i = predicted values u_i

B. Root relative equared error (RRSE)

If relative squared error takes the total squared error and normalizes it by dividing by the total squared error of the simple predictor.

$$\epsilon_i = \sqrt{\frac{\sum_{i=1}^{n}(u_{ij} - x_j)^2}{\sum_{i=1}^{n}(x_j - \bar{x})^2}}$$

U_{ij} is the value predicted by the individual model i the record j.
x_j is target value.

$$\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j$$

IV. EMPIRICAL INVESTIGATION

The Models used are zero R, M5P, SMO Regression and CV Model selection criteria is also used for above used three models. We took Intra day data of date and time, minimum temperature, maximum temperature, temperature, wind speed and visibility of India from 01-01-2017 to 01-01-2019. Wind speed is dependent variable and all other are variables i.e. Date and time, minimum temperature, maximum temperature, temperature, visible are independent variables. By using WEKA software we fitted zero R, SMO Reg, M5P and CV selection criteria is used for all the above 3 models.

The summary and output of SMO regression model for data is correlation coefficient, mean absolute error, root mean squared error, relative absolute error, root relative square error is as follows:

Table 1

Correlation coefficient	0.453
Mean absolute error (MAE)	1.5805
Root mean squared error (RMSE)	2.0021
Relative absolute error (RAE)	87.2273%
Root relative squared error (RRSE)	89.133%

M5P results using WEKA classifier by using wind speed as dependent variable, and date and time, minimum temperature, maximum temperature, temperature and visibility as independent variables is as follows:

Table 2

Correlation coefficient	0.0176
MAE	1.8087
RMSE	2.2448
RAE	99.8205%
RRSE	99.9403%

Correlation coefficient, mean absolute error, root mean squared error, relative absolute error and root relative squared error for zero regression model for atmospheric data from 01-01-2017 to 01-01-2019 (day wise data) is as follows:

Table 3

Correlation coefficient	-0.076
MAE	1.8119
RMSE	2.2462
RAE	100%
RRSE	100%

By using CV parameters criteria, we performed three models SMO reg, Zero R and M5P models. There correlation coefficient, mean absolute error, root mean square error, relative average error, root relative squared error and are listed as below.

Table 4 **CV Parameters estimation**

Model Measures	Zero R	SMO Reg	M5P
CC	-0.076	0.453	0.0176
MAE	1.8119	1.5804	1.8087
RMSE	2.2462	2.0021	2.2448
RAE	100%	87.2248%	99.8205%
RRSE	100%	89.1321%	99.403%

V. SUMMARY AND CONCLUSIONS

For intraday atmospheric data of 01-01-2017 to 01-01-2019, variables used are wind speed as dependent variable and minimum temperature, maximum temperature, temperature, time & date and visibility are independent variables. In this paper M5P, zero regression, SMO Regression models and also CV parameter selection criteria used for M5P, Zero regression and SMO regression models. By using root mean square error criteria and Rank according to their error criteria for selecting the best model.

Table 4

Model	RMSE	Rank
M5P	2.2448	3
Zero R	2.2462	2
SMO R	2.0021	1
CVPSC M5P	2.2448	3
CVPSC Zero R	2.2462	2
CVPSC SMO R	2.0021	1

CV parameter selection criteria and normal models gives the same RMSE values. Among three i.e. M5P, zero R and SMO Reg models for data SMO regression is the best model.

REFERENCES

[1] Zeynab Ramedani, Mahmoud Omid and Alireza Keyhani "Modeling solar energy potential in a Tehran province using artificial neural networks", International Journal of Green Energy, **Vol.10, No.4, pp.427-442, 2013.**

[2] M Cucumo, A. De Rosa, V. Ferraro, D. Kaliakatsos, V. Marinelli, "Experimental testing of models for estimation of hourly solar radiation an vertical surfaces Arcavancata. Di Rende", Solar energy, **Vol. 81, No.5, pp. 692-695, 2007.**

[3] Somaieh Ayalvary, Zohreh Jahani, Morteza Babazadeh, "Select the most relevant input parameters using WEKA for models forecast Solar radiation based on Artificial Neural Networks", ACSIJ Advances in Computer Science: an International Journal, **Vol. 4, No.6(18), pp.38-44, 2015.**

[4] T. Khatib, A.Mohamed, and K. Sopian, "A review of solar energy modeling techniques", Renewable and Sustainable Energy Reviews, **Vol. 16 No.5, pp. 2864-2869, 2012.**

[5] M. A. Abdul Azeez, Artificial Neural Network Estimation of Global Solar Radiation Using Meteorological Parameters in Gusau, Nigeria, Archives of Applied Science Research, **Vol.3, No.2, pp. 586-595, 2011.**

[6] P. Usha Sri and B.Narasimha Swamy, Wireless Atmospheric Data Logger for a Sensor Network, International Journal of Computer Sciences and Engineering, **Vol.3, No.9, PP. 157-161,2015.**

[7] K.V. Shende, V.S.Shirsat , M. R. Ramesh Kumar , K.V. Kale, Artificial Neural Network Model for Prediction of Latent Heat Flux over Bay of Bengal, International Journal of Computer Sciences and Engineering, **Vol.7, No-5, PP.901-905, 2019**

AUTHOR'S PROFILE

Dr. B. Hari Mallikarjuna Reddy is working as an academic Consultant in Dept. Of Statistics, Sri Venkateswara University, Tirupati, A.P, India. He has 13 years of Teaching and Research Experience. He was Published 15 Research papers and published a Book. He was attended and Presented 40 Research Papers in Various National and International Conferences, Seminars and Symposiums in all over India. He also attended 30 Work Shops. He was organized a National Seminar and achieved a National level Award for his research contribution.

Dr. S. Venkatramana Reddy is working as an Associated Professor in the Dept. of Physics, S.V.University and Tirupati. He has 14 years' experience in teaching and research. He was published 110 research papers in various in internationally reputed Journals. He attended and presented 110 papers in various National and International Conferences. He was won Best Paper Award presenting a paper in National Conference. He was awarded PhD's 8 Research scholars under Supervision.

Dr.B. Sarojamma is working as an Associate Professor in Dept. of Statistics, Sri Venkateswara University, Tirupati, A.P, India. She has 14 years of Teaching and Research Experience. She Published 88 Research Articles and 2 Books. She Presented 130 Research Papers in Various National and International Conferences, Seminars and Symposiums in all over India. She attended 57 Woks hops. She organized 4 National Seminars and got One National level award For His Research Contribution. She Guided 5 Ph.D and 2 M.Phil Degrees.

