International Journal of Plant & Soil Science

33(22): 31-38, 2021; Article no.IJPSS.76121
ISSN: 2320-7035

Comparison Statistical Rice Yield Prediction with Multiple Weather Parameters

T. Thurkkaivel[1], G. A. Dheebakaran[1*], V. Geethalakshmi[2], S. G. Patil[3] and K. Bhuvaneshwari[2]

[1]Agro Climate Research Centre, Tamil Nadu Agricultural University, Coimbatore, Tamli Nadu, India.
[2]Directorate Crop Management, Tamil Nadu Agricultural University, Coimbatore, Tamli Nadu, India.
[3]Department of PS& IT, AEC&RI, Tamil Nadu Agricultural University, Coimbatore, Tamli Nadu, India.

Authors' contributions

This work was carried out in collaboration among all authors. Author TT carried out the study of statistical yield forecasting analysis wrote the protocol and wrote the first draft of the manuscript. Author s GAD and Author KB managed the analyses and carried out the correction of "R" language script codes which can be used for the study. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/IJPSS/2021/v33i2230680
Editor(s):
(1) Dr. Hon H. Ho, State University of New York, USA.
Reviewers:
(1) Oyelere Elizabeth Adenike, Ahmadu Bello University, Nigeria.
(2) Prabhuling Tevari, University of Agricultural Sciences, India.
(3) Swadhin Behera, Japan Agency for Marine-Earth Science and Technology, Japan.
Complete Peer review History: https://www.sdiarticle4.com/review-history/76121

Original Research Article

Received 13 August 2021
Accepted 27 October 2021
Published 28 October 2021

ABSTRACT

Advance knowledge of harvestable products, especially essential food crops such as rice, wheat, maize, and pulses, would allow policymakers and traders to plan procurement, processing, pricing, marketing, and related infrastructure and procedures. There are many statistical models are being used for the yield prediction with different weather parameter combinations. The performance of these models are dependent on the location's weather input and its accuracy. In this context, a study was conducted at Agro Climate Research Centre, Tamil Nadu Agricultural University, Coimbatore during *Kharif* (2020) season to compare the performance of four multivariate weather-based models viz., SMLR, LASSO, ENET and Bayesian models for the rice yield prediction at Tanjore district of Tamil Nadu State with Tmax, Tmin, Mean RH, WS, SSH, EVP and RF. The results indicated that the R^2, RMSE, and nRMSE values of the above models were ranged between 0.54 to 0.79 per cent, 149 to 398 kg/ha, 4.0 to 10.6 per cent, respectively. The study concluded that the Bayesian model was found to be more reliable followed by LASSO and ENET. In addition, it

**Corresponding author: E-mail: gadheebakaran@tnau.ac.in, AGRIVELU10@hotmail.com;*

was found that the Bayesian model could perform better even with limited weather parameters and detention of wind speed, sunshine hours and evaporation data would not affect the model performance. It is concluded that Bayesian model may be a better option for rice yield forecasting in Thanjavur districts of Tamil Nadu.

Keywords: Rice yield forecast; statistical models; LASSO; SMLR; ENET; Bayesian.

1. INTRODUCTION

Rice is one of important food crop of the world, being cultivated in more than 100 countries. The tropical weather condition is much preferred by the rice crop as it requires temperature around 30°C and good rainfall. The seasonal variation have good influence on the crop as the long day length and high temperature prevailed during *kharif* shortens the duration of the crop, whereas the short day and lower temperature prevailed during rabi leads to better net photosynthesis and resulted with higher yield in *rabi* crop [1]. The weather parameters such as rainfall, maximum and minimum temperature, relative humidity, evaporation, sunshine hour etc., had markable impact on plant growth and yield [2]. The yield forecasting of the rice and other crops are being issued regularly by government and non-government agencies to ensure the national food security, making decision on crop insurance, import and export plans and subsidies. There are many statistical and crop simulation models are available for crop yield prediction and their performance is vary with input requirements, the capacity of model to perform under various environmental conditions, cost-effectiveness, and level of analytical and statistical experience.

Study on wheat crop yield forecast of nine districts in eastern Uttar Pradesh with statiscal model showed less Root Mean Square Error (RMSE, ± 12%) and coefficient of determination (R^2, 51% and 92%) [3]. Grade [2] inferred that the SMLR was performed better than other statistical models based on the adjusted R^2 value (> 0.7) in wheat crop yield prediction. In the west coast region of India, the Least Absolute Shrinkage and Selection Operator (LASSO) model was found to be best fit for rice yield prediction, based on the R^2, RMSE and nRMSE [4]. In this context, a study was performed at Agro Climate Research Centre, Tamil Nadu Agricultural University, Coimbatore during *Kharif* (2020) season to compare the performance of four multivariate weather-based models viz., SMLR, LASSO, ENET and Bayesian models for the rice yield prediction at Tanjore district of Tamil Nadu State with Tmax, Tmin, Mean RH, WS, SSH, EVP and RF and a part of results are discussed in this paper.

2. MATERIALS AND METHODS

Study Area: The study was taken for the rice crop yield forecast during the *kharif (2020)* season for Tanjore district, which is popularly known as “Rice Bowl of Tamil Nadu”.

Data source: The time series of crop yield data of the Tanjore district of Tamil Nadu for the 29 years (1991 to 2020) was obtained from the Crop and Season report, Department of Economics and Statistics, Chennai, and the past two years crop yield data were collected from the farmers field and Agricultural Office. Daily weather data viz., Tmax, Tmin, RH, WS, SSH, EVP and RF were obtained from the Tamil Nadu Rice Research Institute, Aduthurai, Tamil Nadu. About 85 per cent of the observed data was used for the validation and 15 per cent of data was used for the calibration purpose.

Calculation of the weather indices: Both weighted and unweighted weather indices had been calculated as below for this study and the combinations are depicted in Table.1 and Table 2.

Unweight weather indices

$$Zij = \sum_{w=1}^{m} Xiw \qquad Zii'j = \sum_{w=1}^{m} XiwXi'w$$

Weighted weather indices

$$Zij = \sum_{w=1}^{m} r_{iw}^{j} Xiw \qquad Zii'j = \sum_{w=1}^{m} r_{ii'm}^{j} XiwXi'w$$

Where, m- Week of forecast , Xiw/Xii’- Value of i^{th}/i^{th} weather variable understudy in w^{th}week ,

$r^j_{iw}/r^j_{ii'w}$ – Correlation coefficient of de-trended yield with i^{th} weather variable/product of i^{th} and i'-th weather variables in w^{th} week

Detrending of crop yield: Detrending of yield was done to reduce the nonlinear and non-stationary trend that would cause fluctuation in yield prediction. This trend has to be removed before the computation of basic correlation function in order to improve the performance of the model [5-6]. The simple linear regression model used for the detrending of crop yield was

$Yt = \beta o+\beta 1t$, where Yt - crop yield at given time, βo & β1– Coefficients.

Statistical yield forecasting techniques: In this present investigation, four different linear regression models such as LASSO, ENET, SMLR and Bayesian were used for rice yield prediction and are detailed below.

Stepwise Multiple Linear Regression (SMLR): Stepwise regression is a type of multiple regression that allows to choose the independent variables that will give the greatest prediction with the fewest number of variables. It allows the user to solve a series of one or more multiple linear regression problems using the least square method in a stepwise manner. At each phase of the analysis, a variable is added or eliminated resulting in the biggest error in the sum of squares. Multiple linear regression is an approach used for the development of calibration models. However, it is not always successful when applied to datasets with independent variables. Stepwise multiple linear regression is a procedure that takes into account the feature selection of a linear model. It provided good results in large datasets [3,4,7]. Sometime SMLR is not recommended for the prediction of crop yield because of biased regression coefficients and it removed some variables which are considered as important.

Least Absolute Shrinkage Regression Operator (LASSO): It overcome the drawbacks of ordinary least square (OLS) and ridge regression, through various penalties and retains all predictors. The LASSO model is a regression analysis that does both variable selection and regularisation to improve the statistical model's prediction accuracy and interpretability [6,8,9]. LASSO eliminators are utilised for a consistent regression coefficient and automatic variable selection. LASSO regression produces simpler and more interpretable models that incorporate only a reduced set of predictors.

$$L1= \sum(\hat{Y}i - Yi)^2 + \lambda \sum |\beta|$$

where, y is the independent variable, β is the corresponding coefficient and λ is the L1 norm penalty.

Elastic Net (ENET): It combines both LASSO and RIDGE *i.e.,* penalized with both the L_1 and L_2 norms that effectively shrink coefficients (like in ridge regression) and set some coefficients to zero (like in LASSO). ENET reduces the impact of different features while not eliminating all of the features. [8,6,4,10].

$$L = \sum(\hat{Y}_i - Y_i)^2 + \lambda \sum \beta^2 + \lambda \sum |\beta|$$

Where, y is the independent variable; β is the corresponding coefficient and λ is the L_1 norm penalty.

In ENET model, alpha level fixed at 0.5 whereas alpha <0.5 will have heavier ridge penalty and alpha is >0.5 will have a heavier LASSO penalty. In the present study, the "glmnet" packages used for implementing the LASSO and ENET in R software v.4.1.0 [8, 10].

Bayesian model: This method is known as Bayesian because it is based on Bayes' theorem. It provides true probabilities to quantify the uncertainty about a given hypothesis, but it necessitates the use of a prior belief about how likely this hypothesis is true, known as prior, in order to derive the probability of this hypothesis after seeing the data, known as posterior probability. Bayesian inference, on the other hand, is built on the ability to describe parameter uncertainty using probability theory. It provide a posterior probability distribution over all potential parameter values based on the model and observed data instead of point estimates. It can quickly construct probabilistic statements with the posterior distribution [11, 12]. In this present study, "rstanarm (used the function stan_glm)" packages has been used to predict the rice crop yield in R software v.4.1.0

Testing the model performance:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Oi - Mi)^2}$$

$$R^2 = \left(\frac{1/n\ \sum_{i=1}^{n}(Mi - \bar{M})\,(Oi - \bar{O})}{\sigma_M \sigma_O}\right)^2$$

$$\mathbf{nRMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Oi - Mi)^2} \times \frac{100}{\bar{O}}$$

M_i – model output, $\bar{M}$ and σ_M mean and standard deviation of model output, respectively,

O_i –observation, $\bar{o}$ and σ_o – mean and standard deviation of observation respectively. R^2 value close to 1 and RMSE close to 0 indicates the better model performance.

Percentage of Deviation **=** (Pi-Oi)/Oi ×100,where, Pi-predicted yield, Oi-observed yield.

Model run: The above four models were used to predict the rice crop yield of Thanjavur district, Tamil Nadu at pre-harvesting stage with long term (1991 to 2020) crop yield data as well as daily weather data of seven parameters *viz*., rainfall (RF), maximum temperature (Tmax), minimum temperature (Tmin), average relative humidity (RH), wind speed (WS), sunshine hours (SSH) and pan evaporation (EVP) obtained during 34^{th} to 51^{th} standard meteorological weeks (MSW). In view of identifying the newly incorporated WS, SSH and EVP parameters, each one run was made without any one of these three parameters. Hence, totally four run were performed in each model.

Model Performance: The model performances were computed with the nRMSE (%) values as mentioned by Jamison *et al.,* [13] and categorized as Excellent (<10%), Good (10 - 20%), Fair (20-30%) and Poor (>30%).

3. RESULT AND DISCUSSION

The results obtained from the comparative study on rice crop yield forecast with four different models and four different run at pre-harvesting stages is discussed here. Rice crop yield prediction equation are depicted in Table 3, 4 & 5 for SMLR, LASSO and ENET, respectively. Among the four models compared, highest coefficient of determination (R^2) was observed in Bayesian model and the least was observed in SMLR model in all the four cases of varying weather inputs viz., all, without WS, without SSH and without EVP (Fig. 1). The R^2 value of rice yield forecast at Tanjore district with LASSO and ENET model were comparatively on par to each other and much better than the SMLR yield forecast for Tanjore. It was observed from the results that the model had varying in response to addition or detaining of weather parameters. Detaining the EVP as input invariably reduced the performance of all the models and all other weather parameters. Detention of WS had reduced the performance of LASSO and ENET, whereas the Bayesian and SMLR did not express any reduction in performance. The SSH did not prove its influence in any of the model performance for rice yield forecast of Tanjore District.

The RMSE and nRMSE values were least (4.0% & 149 kg/ha) in Bayesian model and highest (10.2% & 382.6 kg/ha) in LASSO model followed by ENET and SMLR (Fig. 2 and Fig. 3). Detention of EVP increased the RMSE and nRMSE values in all three models except SMLR. Detention of SSH and WS did not have much influence on the performance of LASSO and ENET Models but for Bayesian. Another interesting note that the detention any of three weather parameter viz., WS, SSH and EVP gave positive influence in Bayesian model performance than inclusion of all the parameters.

Table 1. Combination of unweighted weather data

Parameter	Tmax	Tmin	RH	WS	SSH	EVP	RF
T max	Z10						
T min	Z120	Z20					
RH	Z130	Z230	Z30				
WS	Z140	Z240	Z340	Z40			
SSH	Z150	Z250	Z350	Z450	Z50		
EVP	Z160	Z260	Z360	Z460	Z560	Z60	
RF	Z170	Z270	Z370	Z470	Z570	Z670	Z70

Table 2. Combination of unweighted weather data

Parameter	Tmax	Tmin	RH	WS	SSH	EVP	RF
T max	Z11						
T min	Z121	Z21					
RH	Z131	Z231	Z31				
WS	Z141	Z241	Z341	Z41			
SSH	Z151	Z251	Z351	Z451	Z51		
EVP	Z161	Z261	Z361	Z461	Z561	Z61	
RF	Z171	Z271	Z371	Z471	Z571	Z671	Z71

Table 3. Prediction equation for SMLR model

Place	Prediction Equation
Actual	Y=3758.132+Z671*1.982
Without W.S	Y=3758.132+Z671*1.982
Without SSH	Y= 3758.132+Z671*1.982
Without EVP	Y = 4062.165+Z341*0.775+Z371*0.93

Table 4. Prediction equation for LASSO model

Place	Prediction Equation
Actual	Y=5630.5695 + Z21*57.021 + Z121*0.5410+ Z141*0.9819 + Z461*0.1035 + Z671*1.4978
Without W.S	Y=5874.9579+Z21*117.2347+Z671*1.0597
Without SSH	Y=5630.5695 + Z21*57.0216 + Z121*0.5410 + Z141*0.9819 + Z461*0.1035 + Z671*1.4978
Without EVP	Y=6289.9032 + Z21*6.5902 + Z141*1.4229 + Z171*0.1835 + Z341*0.1139 + Z471*0.0852

Table 5. Prediction equation for ENET model

Place	Prediction Equation
Actual	Y=5804.2321+Z21*67.5676+Z41*8.8756+Z121*0.5385+Z141*0.5154+ Z241*0.2167+ Z461*0.6651+Z671*1.3843
Without W.S	Y=6683.4941+Time*0.1105+Z11*1.2754+Z21*123.5240+Z121*0.3894+ Z671*1.5212
Without SSH	Y=5841.2963+Z21*66.1765+Z41*9.3916+Z121*5797+Z141*0.5284+Z241*0.2103+ Z461*0.6091
Without EVP	Y=6308.7136+Z21*31.9184+Z41*4.6961+Z71*0.4081+Z121*1.2950+Z141*0.8266+Z171*0.0836+Z241*0.3835+Z271*0.0700+Z341*0.1466+ Z471*0.2281

Performance of models were compared and expressed in Table 6. Among the four models, the Bayesian model performed Excellent for the rice yield forecast of Tanjore district, both with and without WS or EVP or SSH. Detention of WS improved the performance of LASSO and ENET models in rice yield prediction, whereas detention of SSH and EVP did not have much change on these model performances.

The superior performance of Bayesian model with discriminant analysis techniques is well supported by Vandita Kumari *et al.*, [14]. Similar better performance of LASSO and ENET over SMLR was well supported by Aravind [10] in wheat yield Sridhara *et al.,* [6] in sorghum which was attributed to reduction in over fitting through penalising of the regression coefficient.

Table 6. Model performance

Place	Actual	Without WS	Without SSH	Without EVP
SMLR	Excellent	Excellent	Excellent	Excellent
LASSO	Good	Excellent	Good	Good
ENET	Good	Excellent	Good	Good
Bayesian	Excellent	Excellent	Excellent	Excellent

Performance of statistical models in rice yield forecasting. (Figs.1,2 and 3)

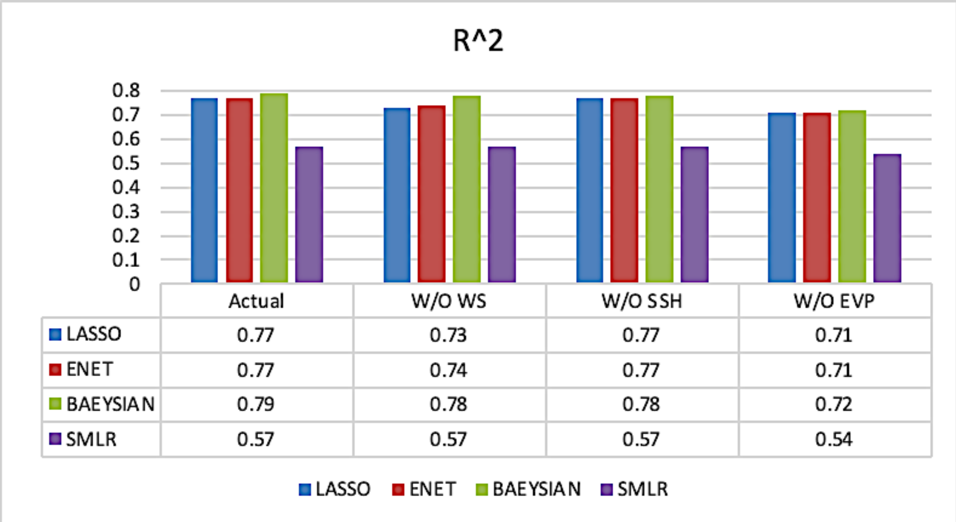

Fig. 1. Coefficient of Determination (R^2) of statistical models for rice yield forecast

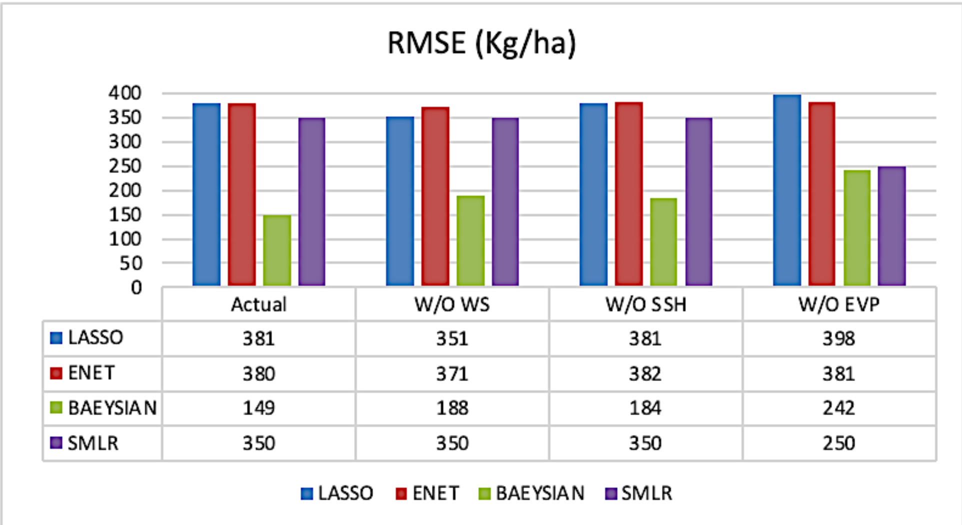

Fig. 2. Statistical models performance with RMSE (kg/ha) for rice yield forecast

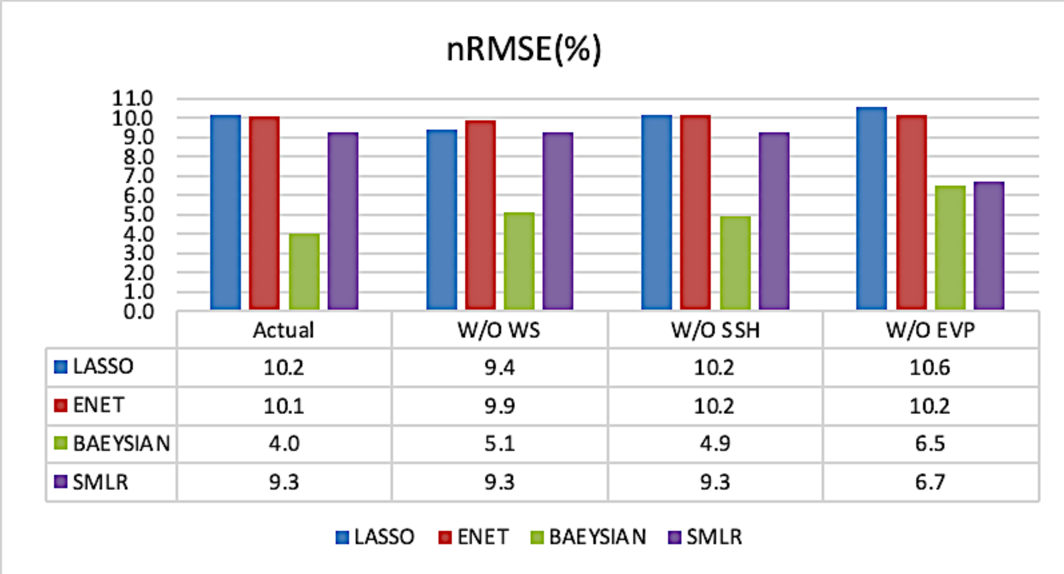

Fig. 3. Statistical models performance with nRMSE (%) for rice yield forecast

4. CONCLUSION

Comparison study of four statistical models' performance for the prediction of *kharif (2020)* season rice yield could be concluded that the Bayesian model was found to be more reliable followed by LASSO and ENET. In addition, it was found that the Bayesian model could perform better even with limited weather parameters and detention of wind speed, sunshine hours and evaporation data would not affect the model performance. It is concluded that Bayesian model may be a better option for rice yield forecasting in Thanjavur districts of Tamil Nadu.

ACKNOWLEDGEMENT

Authors acknowledge the team of scientist in IMD sponsored FASAL scheme operated at Agro Climate Research Centre for sparing methodology and data set.

COMPETING INTEREST

Authors of this article have declared that no competing interests exist.

REFERENCES

1. Mahi GS, Kingra PK. Fundamentals of Agrometeorology and climate change, Kalyani publisher. 2019:135-138.
2. Garde YA, Dhekale BS, Singh S. Different approaches on pre harvest forecasting of wheat yield. Journal of Applied and Natural Science. 2015;7(2):839-843.
3. Singh RS, Patel C, Yadav MK, Singh KK. Yield forecasting of rice and wheat crops for eastern Uttar Pradesh. Journal of Agrometeorology. 2014;16(2):199.
4. Das B, Nair B, Reddy VK, Venkatesh P. Evaluation of multiple linear, neural network and penalised regression models for prediction of rice yield based on weather parameters for west coast of India. International Journal of biometeorology. 2018;62(10):1809-1822.
5. Wu Z, Huang NE. Long SR, Peng CK. On the trend, detrending, and variability of nonlinear and nonstationary time series. Proceedings of the National Academy of Sciences. 2007;104(38):14889-14894.
6. Sridhara, S, Ramesh N, Gopakkali P, Das B, Venkatappa SD, Sanjivaiah S. H, Elansary H. O.Weather-based neural network, stepwise linear and sparse regression approach for rabi sorghum yield forecasting of Karnataka, India. Agronomy. 2020;10(11):1645.
7. Keong YK, Keng W. M.Statistical modeling of weather-based yield forecasting for young mature oil palm. APCBEE Procedia. 2012;4:58-65.
8. Friedman J, Hastie T, Tibshirani R. glmnet: Lasso and elastic-net regularized generalized linear models. R package version. 2009;1(4):1-24.
9. Kumar S, Attri SD, Singh KK. Comparison of Lasso and stepwise regression

technique for wheat yield prediction. Journal of Agrometeorology. 2019;21(2): 188-192.
10. Arvind KS. Multi stage wheat yield estimation using weather based models. M.Sc. Thesis, Division of Agricultural Physics ICAR-Indian Agricultural Research Institute, New Delhi, India; 2019.
11. Raftery AE, Madigan D, Hoeting JA. Bayesian model averaging for linear regression models. Journal of the American Statistical Association. 1997; 92(437):179-91.
12. Bernstein G, Sheldon DR. Differentially private Bayesian linear regression. Advances in Neural Information Processing Systems. 2019;32:525-35.
13. Jamieson PD, Porter, JR, Wilson DRA test of the computer simulation model ARCWHEAT1 on wheat crops grown in New Zealand. Field crops research.1991; 27(4):337-350.
14. Vandita Kumari, Kustav Aditya, Kukum Chandra and Amrendar Kumar. Bayesian discriminant function analysis based forecasting of crop yield in Kanpur district of Uttar Pradesh. Journal of Agrometeorology. 2019;21(4) :460 - 467.

Peer-review history:
The peer review history for this paper can be accessed here:
https://www.sdiarticle4.com/review-history/76121