International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2023; SP-8(5): 170-174

https://www.mathsjournal.com
Received: 02-05-2023
Accepted: 01-06-2023

RS Parmar
Professor, College of Agricultural Information Technology, AAU, Anand, Gujarat, India

GJ Kamani
Assistant Professor, College of Agricultural Information Technology, AAU, Anand, Gujarat, India

YR Ghodasara
Professor, College of Agricultural Information Technology, AAU, Anand, Gujarat, India

Corresponding Author:
RS Parmar
Professor, College of Agricultural Information Technology, AAU, Anand, Gujarat, India

Tree-based ensemble models for productivity trend of minor millets

RS Parmar, GJ Kamani and YR Ghodasara

Abstract
India is one of the leading producers of minor millets, and cultivation of these millets has been declining during the last few years. The present investigation was carried out to study the productivity trend of minor millets in India for the period 1990-91 to 2019-20. Estimation of minor millet productivity trends plays a crucial role in agricultural management in India as an agriculture-based economy. Five tree-based ensemble models *viz*., Bagging Decision Stump, Bagging M5P, Bagging Random Forest, Bagging Random Tree, and Bagging REP Tree were studied. The statistically most fitted tree-based ensemble models were selected based on various performance measurement criteria namely MAE, RMSE, RAE, RRSE, and R^2. The Bagging Random Forest has achieved the highest estimation accuracy of 87% as compared with other fitted tree-based ensemble models. It has the lowest MAE of 33.86 and RMSE of 42.64. Thus, the bagging Random Forest model emerged as the best-fitted trend model for the estimation of the productivity trend of minor millets in India.

Keywords: Bagging, decision stump, random forest, random tree, rapture

Introduction
India is an agriculture-based country and its economy mainly depends on agriculture. Millets can be cultivated in various soil and climatic conditions. The growing season of millets in India is between June to November, and the proper soil type for crop growing is well-drained loamy soil. Millets are rain-fed crops and can be grown up with less rainfall. The information on crop productivity trends plays a crucial role in the preparation and allocation of resources for the growth of agricultural areas. Knowledge of crop productivity trends is important to planners and policymakers for making significant decisions (Nath, 2008) [12]. Panse (1964) [13] said, if any crop is explaining a declining trend in productivity, appropriate strategy measures can be initiated if the productivity data on trend are analyzed well in progress. Barman (2002) [2] studied the productivity trend of coconut in Bangladesh using a 3rd-degree polynomial model. Dayal and Shiam (1963) [4] used linear and curvilinear models for calculating the growth rate of wheat productivity in India for the period 1949-50 to 1961-62. Indiradevi *et al.*, (1990) [8] fitted quadratic, exponential, and semi-log models for the estimation of productivity trends of bananas in Kerala for the period 1970-71 to 1986-87. Goswami *et al.*, (2001) [7] used the model $Y_t = Y_0$ (1+r) to estimate the productivity trend of major crops of Maharashtra for the period 1960-61 to 1996-97. Singh (2002) [20] fitted linear and exponential models for the estimation of the productivity trend of major oilseeds in Gujarat. The best models were selected by comparing their coefficient of determination value. Deka and Sarmah (2005) [5] fitted linear, quadratic, and exponential models to estimate the productivity trend of pineapple in Assam. Rajarathinam *et al.*, (2007) [18] used polynomial models to estimate the productivity trend of sorghum for the period 1979-80 to 1996-97. Parmar *et al.*, (2016) [16] found nonparametric regression with jump point emerged as the best-fitted trend for the area, and production of the cotton crop, and significant jump-points were observed both in area and in production. Parmar *et al.*, (2017) [15] fitted a linear model to estimate the trend in the area and production of maize crop whereas for productivity nonparametric regression without jump-point emerged as the best-fitted trend function. Parmar *et al.*, (2016) [16] observed nonparametric regression without jump-point was selected as the best-fitted trend function for the area, production, and productivity of paddy crop.

Bootstrap Aggregating (Bagging) is an ensemble machine learning technique to reduce the variance of estimation by generating extra data for training from the experimental dataset. The bagging technique is used to avoid overfitting and improves regression accuracy. Archana *et al*., (2020) [1] used ensemble methods to develop a joint system of crop rotation, crop yield prediction, forecasting, and fertilizer recommendation. Kulkarni *et al*., (2018) [10] used ensemble techniques to classify soil types into recommended crop types Kharif or Rabi based on specific physical and chemical characteristics, average rainfall, and surface temperature. Breiman (2001) [3] proposed random forests, which add a layer of randomness to bagging. Yunous Vagh (2012) [21] studied the impact of rainfall on crop productivity using machine learning techniques. Mucherino *et al*., (2009) [11] examined machine learning algorithms models *viz*., K-means, KNN, ANN, and SVM in estimating the productivity trend. Diriba and Borena (2013) [6] used three machine learning methods *viz*., Random Forest, REP Tree, and J48 in estimating crop productivity trend with a prediction accuracy of 83%. Rani and Vidyavathi (2013) [19] estimated sugarcane productivity trend using decision tree and PCA machine learning models. Parmar *et al*., (2021) [17] studied random forest and linear regression models. The random forest model was better with an estimation accuracy of 87% as compared with the linear regression model. Kamani *et al*., (2021) [9] examined linear regression, multilayer perceptron, SMO Reg, gaussian processes, and additive regression models. The most fitted models were selected based on various performances of fit criteria. The MLP has also achieved the highest direction accuracy of 100% and R^2 of 98% as compared with other fitted models. MLP model has the lowest MAE of 80.24 and RMSE of 98.99. Considering the above findings in mind, the present investigation was undertaken to study tree-based ensemble models for the productivity trend of minor millets in India.

Objective

The objective of the study is to develop and compare different fitted tree-based ensemble models for productivity trend of minor millets in India.

Materials and Methods

The ensemble is one of the most accepted and booming methods in machine learning. Ensemble learning is a learning method that consists of combining multiple machine learning models. A problem in machine learning is that individual models tend to present poorly. The individual models are known as weak learners. Weak learners either have a high bias or high variance. Ensemble learning improves a model's performance. Bagging (Bootstrap aggregating) is used to reduce the variance of weak learners. An effort had been made through tree-based ensemble models to learn the trends in the productivity of minor millets in India. The time-series data on the productivity of the minor millets for the period 1990-91 to 2018-19 have been collected from the reports and the India stat website. Open source data mining tool WEKA was used for this research. The time-series dataset is prepared in an Excel sheet with. CSV extension. The five tree-based ensemble models *viz*., Bagging Decision Stump, Bagging M5P, Bagging Random Forest, Bagging Random Tree, and Bagging REP Tree are analysed and the best-fitted model is selected based on various performance measurement criteria *viz*., MAE, RMSE, RAE, RRSE, and R^2 for estimating the productivity of minor millets in India. Fig. 1 shows a semantic diagram of the bagging process.

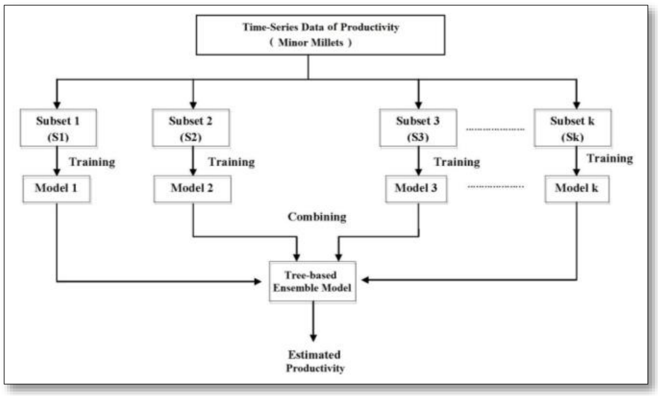

Fig 1: Semantic Diagram of the Bagging Process

Results and Discussion

The Tree-based ensemble models are a powerful machine learning model that combines the estimation from multiple models. An advantage of using Weka for applied machine learning is that IT makes available so many different tree-based ensemble machine learning algorithms. All Tree-based ensemble models are generally driven by three matrices, which comprise the number of independent variables, the form of the regression line, and the type of estimated variable. Five tree-based ensemble models are evaluated namely Bagging Decision Stump, Bagging M5P, Bagging Random Forest, Bagging Random Tree, and Bagging REP Tree. The performance of each fitted tree-based ensemble model is checked in terms of MAE, RMSE, RAE, RRSE, and R^2. The characteristics of fitted tree-based ensemble models in Table 1 show that the Bagging Random Forest model has better performance than other fitted models. In general, it could be noticed that Bagging Random Forest is the best-fitted tree-based model to estimate the time series dataset of Minor Millets in India.

Table 1: Characteristics of Fitted Tree-based Ensemble Models

Tree-based Ensemble Models	Performance Parameters				
	MAE	RMSE	RAE	RRSE	R^2
Bagging Decision Stump	40.48	52.49	41.56%	44.11%	81.0%
Bagging M5P	37.91	49.22	38.93%	41.36%	84.0%
Bagging Random Forest	33.83	42.64	34.74%	35.83%	87.0%
Bagging Random Tree	38.46	46.67	39.49%	39.22%	84.0%
Bagging REP Tree	35.63	47.58	36.58%	39.98%	84.0%

Fig. 1 depicts the estimation accuracy of fitted tree-based ensemble models. Out of five ensemble models used in this study, Bagging Random Forest has better estimation accuracy than other ensemble models with 87.0%, followed by Bagging M5P, Bagging Random Tree, and Bagging REP Tree with 84.0%**.** Bagging Decision Stump has the lowest estimation accuracy of 81.0%.

Fig. 2 demonstrates the error results of fitted tree-based ensemble models. Bagging Random Forest has the lowest MAE of 33.83 and RMSE of 42.64. This pictures minimal error reported during the estimation processes. Bagging Decision Stump has the highest error rate with 40.48 and 52.49 of MAE and RMSE respectively.

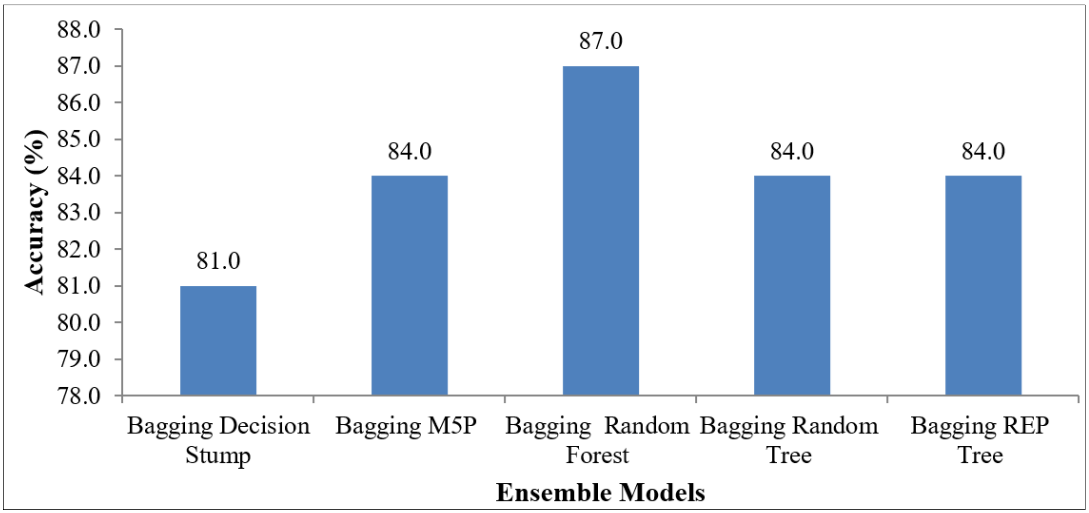

Fig 1: Estimation Accuracy of Fitted Tree-based Ensemble Models

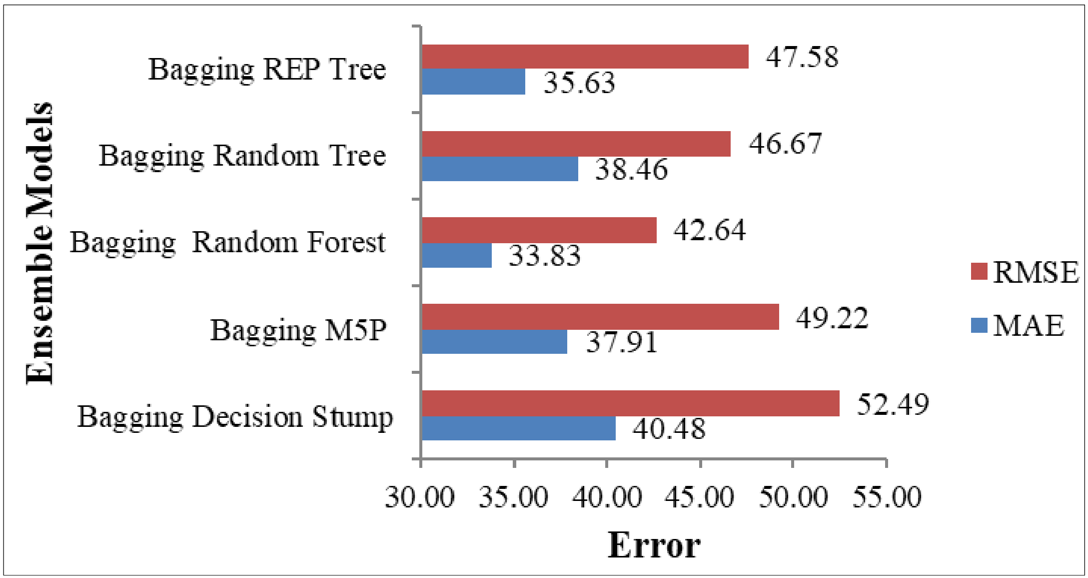

Fig 2: Error Results of Fitted Tree-based Ensemble Models

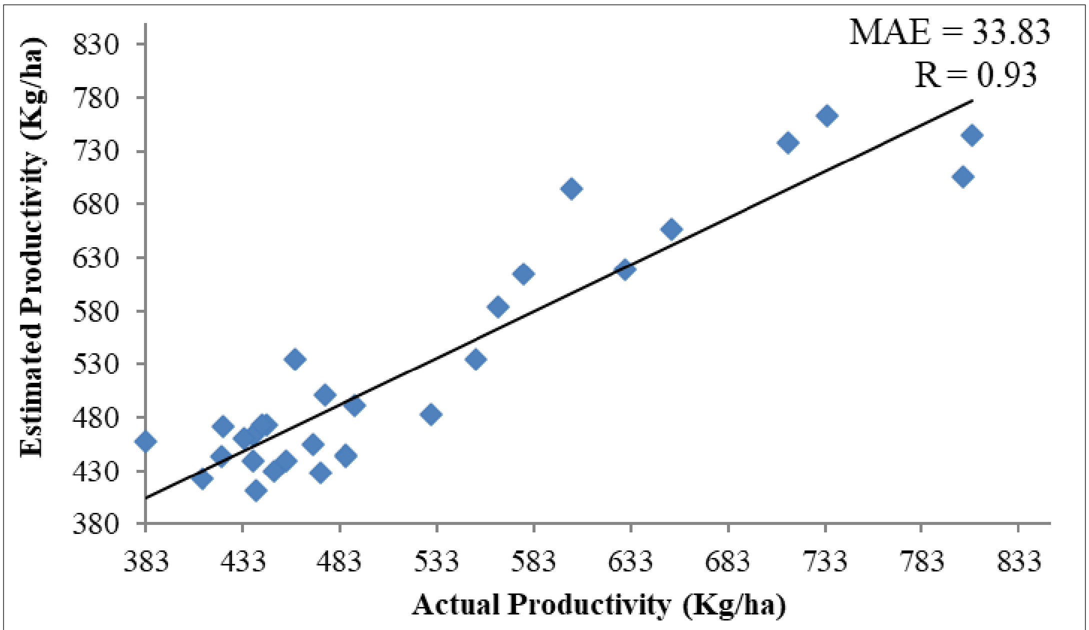

Fig 3: Mean Absolute Error of Best Fitted Bagging Random Forest Model

Fig.3 depicts the MAE and Correlation Coefficient of the fitted Bagging Random Forest model. The low MAE (33.83) and higher R-value (0.93) suggest the fitted tree-based ensemble model is good at the estimation of the productivity trend.

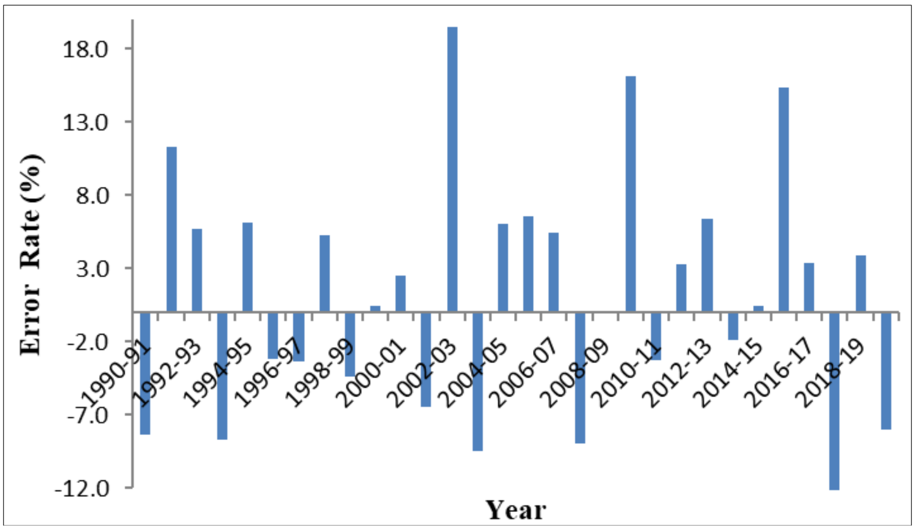

Fig 4: Estimated Productivity Error Rate of Best Fitted Bagging RandomForest Model

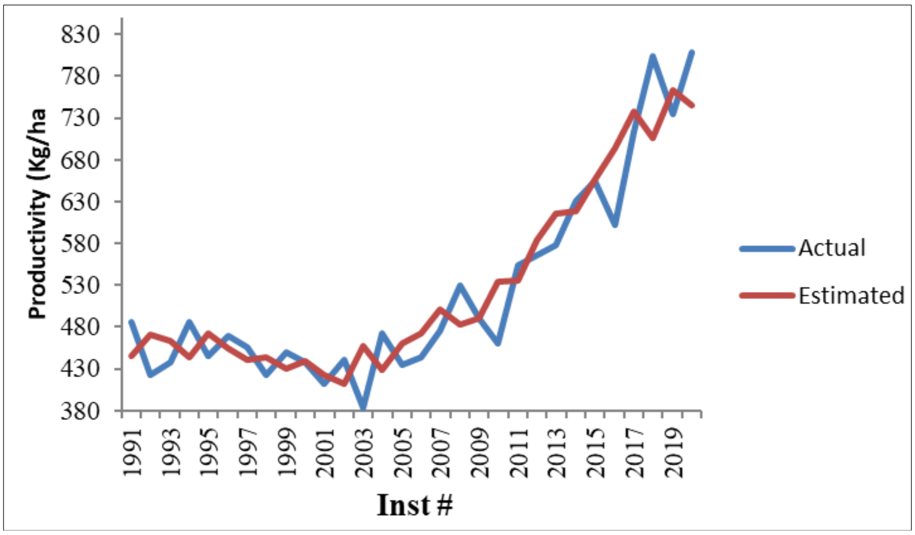

Fig 5: Actual and Estimated Productivity Using Best-Fitted Bagging Random Forest Model

The estimated productivity error rate of the best-fitted Bagging Random Forest model as shown in Fig.4 explicates that the estimated productivity error rates were upward or downward for different years. The estimated productivity error rates were downward by 8.4%, 8.7%, 3.2%, 3.3%, 4.4%, 6.5%, 9.4%, 8.9%, 0.1%, 3.2%, 1.9%, 12.2%, and 8.0% for the years 1990-91, 1993-94, 1995-96, 1996-97, 1998-99, 2001-02, 2003-04, 2007-08, 2008-09, 2010-11, 2013-14, 2017-18 and 2019-20 respectively. But, estimated productivity error rates were upward by 11.3%, 5.7%, 6.2%, 5.3%, 0.4%, 2.5%, 19.5%, 6.1%, 6.6%, 5.4%, 16.2%, 3.3%, 6.4%, 0.4%,15.4%, 3.4%, and 3.9% for the years 1991-92, 1992-93, 1994-95, 1997-98, 1999-2000, 2000-01, 2002-03, 2004-05, 2005-06, 2006-07, 2009-10, 2011-12, 2012-13, 2014-15, 2015-16, 2016-17, and 2018-19 respectively. The estimated productivity error rates ranged from –12.20% to 19.5%.

The actual and estimated Productivity using the best-fitted Bagging Random Forest model on the experiment data set is presented in Fig.5. It is found that the actual and the estimated productivity are close to each other.

Conclusion

Based on all the yardsticks used to measure the estimation of tree-based ensemble models for the productivity trend of minor millets in India, it was found that the Bagging Random Forest model materialized as the best-fitted tree-based trend model for the productivity of minor millets in India by achieving the highest coefficient of determination (R^2) of 87% and lowest MAE of 33.83 as compared with other fitted ensemble models. Hence, it can be concluded that the study helps the researchers in efficient algorithm selection for the productivity trend of minor millets in India. Extension personnel is advised to use the Bagging Random Forest model for estimating the productivity trend of minor millets in India for the farming community.

References

1. Archana K, Saranya KG. Crop Yield Prediction, Forecasting and Fertilizer Recommendation using Voting Based Ensemble Classifier., Seventh Sense Research Group (SSRG) International Journal of Computer Science and Engineering. 2020;7(5):1-4.
2. Barman SC. Production and price of coconut (*Cocus nucifera* L.) and its oil in Bangladesh. Economic Affairs. 2002;47(4):221-227.
3. Breiman. Random forests., Machine Learning. 2001;45(1):5-32.
4. Dayal R, Shiam R. Measurement of growth rates of agricultural production. Agricultural Situation in India. 1963;18(4):177-181.

5. Deka N, Sarmah AK. An analysis of growth trends in area, production and productivity of pineapple in Assam. Economic Affairs. 2005;50(2):110-113.
6. Diriba Z, Borena B. Application of data mining techniques for crop productivity prediction. HiLCoE J Comp. Sci. Tech. 2013;1:151-155.
7. Goswami SN, Khandare NC, Hajare TN, Sen TK. Growth Trend of Principal Crops of Maharashtra: An Analytical Approach. J Maharashtra agric. Univ. 2001;26(1):90-93.
8. Indiradevi P, Thomas EK, Thomas JK. Growth and supply response of banana in Kerala. Agricultural Situation in India. 1990;45(4):239-242.
9. Kamani GJ, Parmar RS, Ghodasara YR. Machine learning models for productivity trend of Rice crop, Guj. J Ext. Edu. 2021;32(2):189-196.
10. Kulkarni JG, Kumar S, Raj RA. Digital soil mapping in a Himalayan watershed using remote sensing and terrain parameters employing artificial neural network model., Environmental earth sciences. 2018;77(5):1-14.
11. Mucherino A, Papajorgji P, Pardolas PM. A survey of data mining techniques applied to agriculture. Operat. Res: An Intl. J. 2009;9(2):121-140.
12. Nath SK. Manual on area and crop production statistics. Central Statistical Organisation; c2008.
13. Panse VG. Yield trends of rice and wheat in first two five-year plans in India. J Ind. Soc. Agril. Statist. 1964;16:1-50.
14. Parmar RS, Chaudhari GB, Parmar DK. Application of Regression Models for Area, Production and Productivity Trends of Maize (*Zea mays*) Crop for Panchmahal Region of Gujarat State, India., Intl. J on Recent and Innovation Trends in Computing and Communication. 2017;5(1):303-308.
15. Parmar RS, Patel HK, Dabhi BK, Vegad NM. Estimating Statistical Models for Area, Production and Productivity Trends of Paddy (*Oryza sativa* L.) Crop for Anand Region of Gujarat State, India. Advances in Life Sciences. 2017;5(16):6025-6031.
16. Parmar RS, Rajarathinam A, Patel HK, Patel KV. Statistical Modeling on Area, Production and Productivity of Cotton (*Gossypium* spp.) Crop for Ahmedabad Region of Gujarat State., J Pure and Applied Microbiology. 2016;10(1):751-759.
17. Parmar RS, Vegad NM, Raj MP. RandomForest and regression analysis for an attitude of research Scholars towards Internet exposure., Guj. J Ext. Edu. 2021;32(1):135-139.
18. Rajarathinam A, Dixit SK, Vaishnav PR. Fitting of Sorghum (*Sorghum bicolor*) yield trends in long-term fertilizer experiment. Crop Res. 2007;34(1, 2, 3):57-63.
19. Rani A, Vidyavathi BM. Ameliorated methodology for the design of sugarcane yield prediction using decision tree. Compusoft. An Int. J Adv. Comp. Tech. 2013;4:1882-1889.
20. Singh OP. An analysis of major oilseed production in Gujarat. Economic Affairs. 2002;47(4):234-243.
21. Yunous Vagh. An Investigation into the Effect of Stochastic Annual Rainfall on Crop Yields in South Western Australia. International Journal of Information and Education Technolog; c2012, 2(3).