RELATIONSHIP BETWEEN PROFILE OF SCHOLARS PURSUING POST GRADUATION IN AGRICULTURAL EXTENSION AND THEIR OVERALL E-EXTENSION EMPLOYABILITY USING REGRESSION MACHINE LEARNING ALGORITHMS

N. M. Vegad[1], R. S. Parmar[2] and Y. R. Ghodasara[3]

1 Assistant Professor, College of Agricultural Information Technology, AAU, Anand - 388110 India
2&3 Professor, College of Agricultural Information Technology, AAU, Anand - 388110 India
Email: rsparmar@aau.in

ABSTRACT

The e-extension services to agriculture extension are a network of institutes that provides more efficient alternative to the traditional extension system for agriculture and its allied sectors. Besides, to achieve another main objective of a study on the relationship between profile of scholars pursuing post graduation in agricultural extension and their overall e-extension employability, five regression algorithms namely, linear regression, multilayer perceptron , support vector machines, k-nearest neighbors and KStar were used in the study. Among these tested algorithms, KStar was better with prediction accuracy of 94 % (R^2), lowest Mean Absolute Error (MAE) of 0.82 and Root Mean Squared Error (RMSE) of 1.95 as compared with other fitted regression algorithms. It was further noticed that, the actual overall e-extension employability and the predicted overall e-extension employability based on testing data set were close to each other and the residual ranged from -6.17 to 4.82.

Keywords*: linear regression, multilayer perceptron, support vector machines, k-nearest neighbors*

INTRODUCTION

E-extension in agriculture is a new area of knowledge emerging out of convergence of ICT and farming techniques (Jagadeeswari *et al.,* 2019). It augments the agricultural value chain through the use of ICT technologies. The agriculture sector is gearing itself to make optimal use of the e-extension to get predictable results (Darji *et al.,* 2017). At the management levels, a number of significant e-extension based initiatives have been taken to provide connectivity to all the societies involved in agricultural education, research and dissemination. Thus it is necessity for every agricultural extension educationists to have employability and enthusiasm in terms of knowledge, attitude and exposure of all those basics and devices which are directly or indirectly required to investigate e-extension for agricultural growth. Regression Machine Learning is a subset of artificial intelligence that is mostly concerned with the expansion of regression algorithms which allow a computer to learn from the secondary data. The term machine learning was first introduced by Samuel Arthur (1959). He said that machine learning enables a machine to automatically learn from secondary data, improve performance from experiences, and predict things without being explicitly programmed. Manish (2009) said that machine learning is a growing technology with the growth of artificial intelligence and database procedures which is used in different business organization to improve the effectiveness and significance of a business process. Machine learning is a multidisciplinary field that combines artificial intelligence, computer science, data mining, mathematics algorithms and statistics (Liao SH 2003). A regression machine learning algorithms in agriculture was projected by Fathima and Geetha (2014). Machine learning technique uses regression algorithms to find out useful information from huge data set, for industry operation (Witten and Eibe 2011). Kumari and Chitra (2013) used support vector machine learning algorithm for predicting diabetes. Their experimental results showed that SVM can be successfully used for predicting diabetes diseases. In this study, five regression algorithms namely linear regression, multilayer perceptron, support vector machines, k-nearest neighbors and KStar were fitted most accurate and effective models. Moreover, regression machine learning algorithms take a data-driven technique to learn useful relationships from experimental data set (Willcock et al., 2018) and provide a best way for improving predictions. In addition, regression algorithms have some individual benefits like; they can model non-linear relationships between multiple data sources (Chlingaryan et al. 2018). The present study will be useful for researchers to know the e-extension employability of students studying in higher agricultural extension education. The study would be helpful in knowing the relationship between profile of

the postgraduate agricultural extension students and their e-extension employability.

OBJECTIVE

To study the relationship between profile of scholars pursuing post graduation in agricultural extension and their overall e-extension employability using regression machine learning algorithms

METHODOLOGY

The present research was conducted in State Agricultural Universities of Gujarat. The ex-post facto research design was use for the investigation. Data were chosen from a random sample of 120 researchers seeking post graduation in agricultural extension in SAUs of Gujarat. The Dataset having 18 feature variables namely academic performance, native place, father's education, mother's education, annual family income, family occupation, involvement in extracurricular activity, medium of education, information collection behavior, attitude towards extension work, library exposure, job preference, achievement motivation, self confidence, interpersonal communication, innovativeness ,professional zeal , willingness to work in rural area. E-extension employability is target variable. The experimental dataset was created in excel sheet with .CSV extension for research by regression machine learning system (Weka). Normalized algorithm was used to normalize the experimental dataset. Selected feature variables by Attribute evaluator namely **"cfsSubsetEval"** and search method namely "**BestFirst**" are academic performance, annual family income, library exposure, information collection behavior, job preference, self confidence, interpersonal communication, willingness to work in rural area. The five regression algorithms namely linear regression, multilayer perceptron , support vector machines, k-nearest neighbors and Kstar were used over the experimental data set. Coefficient of determination (R^2), MAE, RMSE, Relative Absolute Error and Root Relative Squared Error were taken into consideration for comparison each regression algorithm.

RESULTS AND DISCUSSION

A weka is a collection of regression machine learning algorithms for regression analysis. Regression machine learning algorithms are usually driven by the number of input variables, the character of regression line and the type of dependent variables. From Weka, five regression algorithms are evaluated namely linear regression (LR), multilayer perceptron (MLP), support vector machines (SMOReg), k-nearest neighbors (IBK) and KStar (K*). The result of each regression algorithm is checked in terms of R^2, MAE, RMSE, RAE and RRSE**.** The Fig. 1 reveals that the selected attributes have differed distribution range.

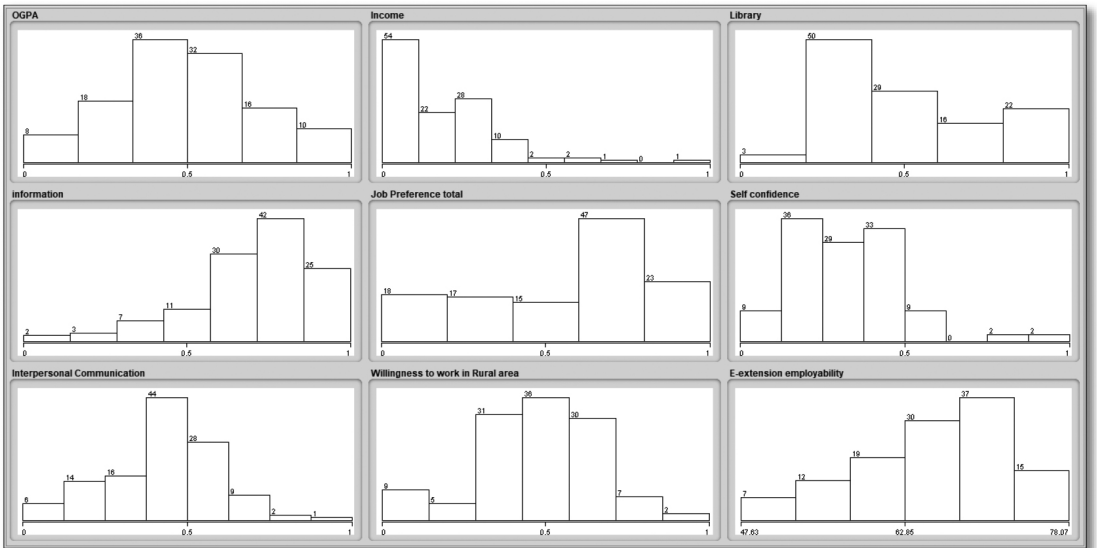

Fig. 1 : Selected Attribute Distributions

The regression analysis of overall e-extension employability in Table 1 indicates that the lazy based algorithms have better performance than function based algorithms**.** In case of function based algorithms, three algorithms are observed namely linear regression, multilayer perceptron and SMOReg. SMOReg has better performance than other fitted algorithms. For lazy based algorithms, two algorithms are examined viz., k-nearest neighbors (IBK) and KStar. The KStar model has better performance than k-nearest neighbors (IBK). In general, it could be observed that, KStar algorithm is better than SMOReg algorithm further underlining that the KStar provides higher accuracy than other regression algorithms.

Table 2: Regression analysis of overall e-extension employability

Parameters	Regression Algorithms				
	Functions Based			Lazy Based	
	Linear Regression	Multilayer Perceptron	SMO Reg	k-Nearest Neighbors	KStar
Mean Absolute Error (MAE)	3.76	3.92	3.00	1.50	**0.82**
Root Mean Squared Error (RMSE)	4.76	4.95	4.13	3.50	**1.95**
Relative Absolute Error (RAE)	57.63 %	60.07 %	45.96 %	22.92 %	**12.58 %**
Root Relative Squared Error (RRSE)	59.53 %	61.71 %	51.46 %	43.68 %	**24.29 %**
Coefficient of Determination (R^2)	67 %	66 %	75 %	82 %	**94 %**

The Fig. 2 demonstrates the coefficient of determination (R^2) of fitted regression algorithms. Out of five algorithms used in this research work, KStar algorithm has better prediction accuracy than other fitted regression algorithms with 94 %, followed by KNN with 82 %. Linear regression has lowest prediction accuracy with 67%.

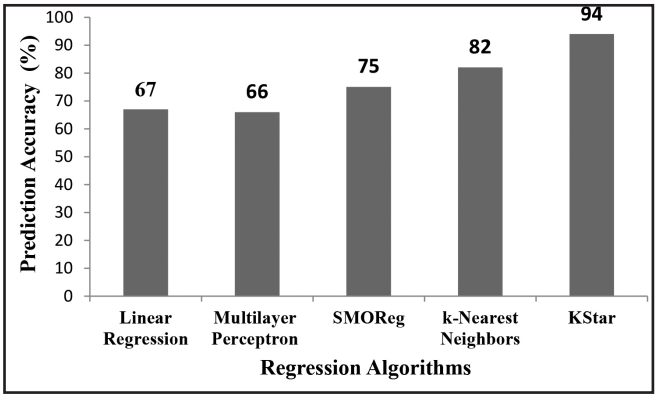

Fig. 2: Coefficient of determination of fitted regression algorithms

The Fig. 3 shows the error results of the different regression algorithms. KStar has lowest Mean Absolute Error (MAE) of 0.82 and Root Mean Squared Error (RMSE) of 1.95. This reveals minimal error reported during the prediction processes. Liner regression has the highest error rate with 3.76 and 4.76 of MAE and RMSE respectively.

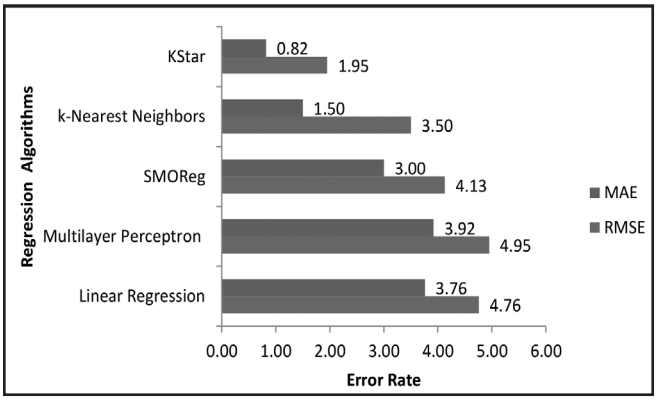

Fig. 3 : Error results of fitted regression algorithms

Table 3: Actual and predicted overall e-extension employability using kstar

Sr. No.	Overall E-Extension Employability			Sr. No.	Overall E-Extension Employability		
	Actual	Predicted	Residual		Actual	Predicted	Residual
1	75.56	75.38	-0.18	13	53.15	53.76	0.61
2	64.43	64.10	-0.33	14	51.96	52.13	0.17
3	69.52	69.34	-0.18	15	47.63	48.31	0.68
4	61.96	66.78	4.82	16	76.13	70.78	-5.35
5	73.59	73.70	0.11	17	71.98	65.81	-6.17
6	66.93	66.88	-0.05	18	72.56	72.56	0.00
7	75.30	75.41	0.11	19	70.10	70.10	0.00
8	69.78	69.60	-0.18	20	72.11	72.11	0.00
9	63.63	63.48	-0.15	21	51.78	51.79	0.01
10	68.28	68.02	-0.26	22	65.72	65.51	-0.21
11	62.67	62.52	-0.15	23	62.50	62.50	0.00
12	63.74	63.74	0.00	24	56.21	56.21	0.00

The predicted overall e-extension employability based on the testing data set is presented in Table 3. The same is demonstrated in Fig.4.

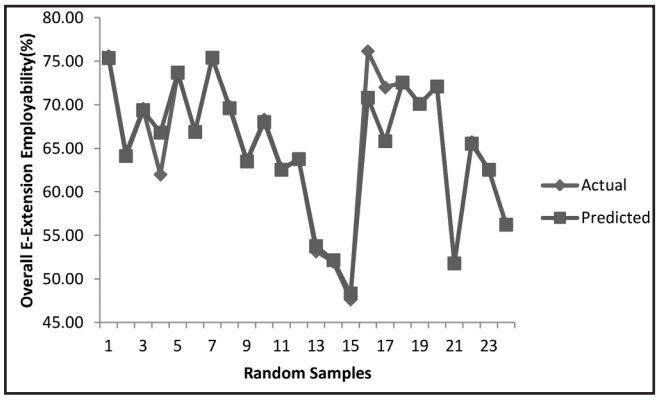

Fig 4 : Actual and predicted overall e-extension employability using Kstar

It is observed that, the actual overall e-extension employability and the predicted overall e-extension employability are closed to each other. The residual ranged from -6.17 to 4.82.

CONCLUSION

A study on the relationship between profile of scholars pursuing post graduation in agricultural extension and their overall e-extension employability is an important dimension of its overall e-extension employability prediction. Five regression algorithms namely, linear regression, multilayer perceptron, support vector machines, k-nearest neighbors and KStar were used in the study as these have been gaining popularity in agricultural extension applications due to its success in prediction. Among these, KStar algorithm was found superior with predictability of 94 % (R^2) , lowest Mean Absolute Error (MAE) of 0.82 and Root Mean Squared Error (RMSE) of 1.95 as compared with other fitted regression algorithms. The study can be further extrapolated and made the model robust by interfacing with GIS for better utilization by the stake holders.

REFERENCES

Chlingaryan, A., Sukkarieh, S. and Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: a review. *Comput. Electron. Agric*. 151, 61–69.

Darji, A. R.. Panchasara B. R. and Vinaya Kumar H.M. (2017). Determinates of attitude of postgraduate research scholars towards the use of computer for their empowerment. *Guj. J. Ext. Edu.,* 28 (2): 277-281.

Fathima, N G and Geetha, R. (2014). Agriculture crop pattern using data mining techniques. Intl.J. Adv. Res. Comp Sci. Soft. Eng., 4: 781-785.

Jagadeeswari, B., Vinaya Kumar H. M., & Patel, J. B. (2019). Attitude of postgraduate students towards research, *Guj. J. Ext. Edu.,* 30(1), 87-89.

Kumari, AV. and Chitra, R. (2013). Classification of Diabetes Disease Using Support Vector Machine. *International Journal of Engineering Research and Applications*(IJERA)., 7(1) : 67:70.

Liao, SH. (2003). Knowledge Management Technologies and applications Literature review from 1995 to 2002. Expert System with Application., 25:155-164

Manish, Jain. (2009). Data Mining: Typical Data Mining Process for Predictive Modeling. BPB Publications. , 235-241

Samuel, Arthur L. (1959). Some Studies in Machine Learning Using the Game of Checkers". *IBM Journal of Research and Development.* 44: 206–226.

Witten, IH.; Eibe, Frank. (2011). Data Mining: Practical Machine Learning Tools and Techniques. Second Edition (Morgan Kaufmann Series in Data Management Systems). , 189-303.

Willcock, S., Hooftman, D.A., Bagstad, K.J., Balbi, S., Marzo, A., Prato, C., Sciandrello, S., Signorello, G. and Voigt, B. (2018). Machine learning for ecosystem services. *Ecosystem services* 33, 165–174.

Received : August 2021 : Accepted : November 2021