International Journal of Soft Computing 12 (2): 96-102, 2017
ISSN: 1816-9503

Prediction of Crop Yield Using Regression Techniques

Aditya Shastry, H.A. Sanjay and E. Bhanusree
Nitte Meenakshi Institute of Technology, Bangalore, India

Abstract: With the emergence of artificial intelligence and computer science, data mining has received an enormous amount of boost. Recently, data mining algorithms have been successfully used in the field of agriculture for predicting the yield of crops. Crop yield prediction involves predicting the yield of crops from available historic data like weather parameters, soil parameters and historic crop yield. Regression is a data mining function that predicts a number. Regression techniques are very useful in predicting the yield of crops. In this study, the focus is on the development of regression techniques in agricultural field. Different regression techniques such as quadratic, pure-quadratic, interactions and polynomial are used for predicting the yields of wheat, maize and cotton crops. Finally regression models are proposed which are able to accurately predict the yields of cotton, maize and wheat. The best regression model is selected based on Root Mean Squared Error (RMSE), R^2 and Mean Percentage Prediction Error (MPPE) values.

Key words: Regression, yield, parameters, model, accuracy

INTRODUCTION

Data mining (the investigation venture of the "Learning Discovery in Databases" procedure), a field at the intersection of software engineering and measurements is the procedure that endeavors to find designs in extensive information sets. Data mining covers concepts of databases, machine learning, statistics and artificial intelligence. The main objective of data mining is to find useful patterns from existing data. The genuine data mining assignment is the semiautomatic or automatic analysis of large amounts of information to extract already unknown fascinating examples, like unusual records, conditions and gatherings of information records which help in providing accurate prediction results using predictive modeling

From the past few years predictive modeling is practiced in only few areas in this competitive world. The big data phenomenon is a tool which is used for data analysis in new applications in order to increase the adoption of predictive models. Predictive models assist decision makers in forming right decisions by making them efficient and effective. The automation of entire decision making process is also possible in some cases. Agricultural system is very complex, since it deals with large data situation which comes from a number of factors. Crop yield prediction is significant for farmers and agriculture related industries.

In this research, regression models for predicting the yield of crops like cotton, wheat and maize depending on soil weather and crop parameters are proposed. Comparison is made between the different regression models based on RMSE, R^2 and MPPE metrics.

Literature review: In this study, research works related to the Data mining techniques for crop yield prediction is discussed. Zaefizadeh *et al.* (2011), 40 genotypes were planted in Ardabil. Multiple Linear Regression (MLR) and Artificial Neural Networks (ANN) were employed for prediction of grain yield. In ANN 15 neurons with one hidden layer was utilized. The activation function and learning method used were hyperbolic tangent function and error propagation respectively. Experimental results established that MLR outperformed ANN.

Sanchez *et al.* (2014), researchers show a correlation among a few strategies (linear and nonlinear) for prediction of crop yield. The comparison is made utilizing the best property subset found in the preparation dataset for every strategy which was distinguished utilizing the percentage split validation and a complete algorithm. To search the optimal attribute subset in training datasets the algorithm uses the oldest samples to build the models. The test datasets is composed of unseen samples where the performance is measured. The most widely recognized information driven method for prediction of crop yield like stepwise linear regression, multiple linear regression, regression trees and neural systems were assessed. The experimentation demonstrates that our quality determination utilizing a

Corresponding Author: Aditya Shastry, Nitte Meenakshi Institute of Technology, Bangalore, India

complete system generously enhances the execution of all the assessed strategies. ANN and M5 acquired the best forecast and, between them, ANN accomplished the lower RRSE, the higher R relationship and the lower RMAE value. By and by, none of the strategies had the capacity to obtain the optimum subset with the training data for all the eight yields. The best method was ANN which accomplished three attribute subsets equal to the optimal and the other two subsets were very close to it. Thus, an attribute subset that can be used permanently in all the years for all the crops is difficult to select. Results obtained from machine-learning methods cannot be straightforwardly connected to an alternate arrangement of harvest databases, due their high dependency of information. The strategy presented in this paper can be extended for a larger number of techniques and crop datasets. A future research focused on finding the best minimal subset of attributes which provides a good yield of predictions on other irrigation zones should be considered.

Zhang *et al*. (2010), researchers consider the linear regression model taking into account that the Ordinary Least Square (OLS) estimation is a generally utilized strategy for prediction of crop yield. Here, autoregressive model performed better than OLS with higher R^2. The research concluded that NDVI and precipitation contributed more to the corn yield in Iowa, excluding temperature.

Zaw and Naing (2009), the researchers consider the Polynomial Regression Model (MPR) in order to predict the rainfall in the region of Myanmar. The authors have created prediction forecast model in view of 15 predictors utilizing second-arrange MPR. As a consequence of a few examinations, the anticipated precipitation sum is near to the genuine qualities. SMR expectation model was created with four indicators. The model results were used in the territories, for example in harvest planning and yield prediction, water administration and repository control. The fundamental point of the improvement of the forecast model is to help in water management and farm management. All possible subsets of predictors have been examined in the implementation of multiple polynomial regression which utilizes only 2006 test data. Authors demonstrate that MPR performs better than MLR.

Qaddoum and Hines (2012), researchers perform an expansion to the conventional regression neural networks. Conformal Prediction (CP) framework is built for predicting the tomato yield in a greenhouse taking into consideration Vapor Pressure Deficit (VPD), CO_2, radiation and temperature. About 60,000 records were used in this process.

Ramasubramanian researchers discuss the different forecasting approaches in agriculture including regression models, time series models and probabilistic models. In regression model there are three models that are multiple linear regression models for forecasting crop yields; Weather indices based multiple linear regression model for crop pest count and logistic regression model for forewarning qualitative response variables. In time series model two models used are exponential smoothing model for production of crops and auto regressive integrated moving average models. In probabilistic model, Markov Chain Model is utilized for crop yield forecasting.

MATERIALS AND METHODS

Experimental data with parameters: Data sources along with the input parameters considered for the prediction of wheat, maize and cotton yields is elaborated in this section. To perform the analysis, it is important to gather the datasets from various sources. The information has to be in a specific form which has to be altered again based upon the attributes. In order to give the input to the regression model following data sets have been considered. Datasets gathered is for three harvests, wheat, cotton and maize. Entire datasets comprising of number of records with different attributes which influences the product yield are presented.

Maize data set was collected by Carberry and Abrecht (1991) having 78 records. Wheat data set was collected by Asseng *et al*. (1998) with a total of 50 records. Cotton Data Set was collected by Hearn (1994) having 123 records. The attributes play a very important role in crop yield prediction but not all attributes give proper predicted values. Attributes which influence the harvest yield the most are chosen. The attribute selection assumes essential part because relying upon the chosen attributes the yield of the crop can be predicted more accurately. There are only some specific attributes which help to process the yield output. Table 1 lists the attributes that affect the yield of the respective crops the most.

Regression model design for crop yield prediction: In this study, the general design followed for developing the regression models for crop yield prediction is discussed. Regression analysis is majorly used for prediction purposes as it provides predicted entity as a function of the dependent entities. In certain cases, it gives the relationships between independent and dependent variables (Alan, 1993). The steps to develop a regression model for prediction of crop yield.

Table 1: Attributes for crop yield prediction

Attributes for wheat yields	Attributes for maize yields	Attributes for cotton yields
Biomass (dry water content (kg/ha))	ESW	dw_total (total dry weight requirement)
Grain_protein (amount of proteins (%))	Biomass	NUptake (Nitrogen uptake)
Grain size (size of grain content in grams)	Rainfall (mm)	bolls_sc (No. of cotton bolls per section)
ESW (extractable soil water (mm)	Radiation (solar radiation (mJ/m^2))	lai_max (Max leaf area index (m^2))
		ESW
		Rainfall
		Radiation
		Maximum temperature (°)
		Minimum temperature

Algorithm; steps for crop yield prediction using regression:

Input: Experimental data set of weather data, crop data and soil data
Output: Predicted crop yield for the experimental dataset.
Method:

Step1: Gather, format and organize the information: Only raw information is insufficient to work with the model. The information must be gathered, sort out as per the necessity and organization it in such a path, that suitable results are obtained. While redoing, more important information can be included

Step 2: Separate information into testing and training sets: the data information needs to go be partitioned into two sets. Training set will have greatest rate of the information in order to train most of the examples to make the yield. Approximately 70% of the samples are collected under training set. Testing set uses the remaining measure of the information to test how the system is performing.

Step 3: Apply regression on trained sets: the model system relies on upon how complex the issue is and the structure likewise must be chosen with the need. While altering, the construction modeling and structure can be adjusted.

Step 4: Calculate the RMSE, R^2 statistics and MPPE values for the different models.

Step 5: Apply the trained regression model on test set and again calculate the RMSE, R^2 statistics and MPPE values. Compare the values with different models of regression models. The model which gives the lowest RMSE and MPPE values and highest R^2 statistics values is considered to be the best model for crop yield prediction.

Figure 1 shows the flow chart for regression methodology used for predicting the crop yield.

Regression model for maize yield prediction: Here, the maize data set consists of a total of 78 records. Out of which 55 records are considered as training set and remaining 23 records as test set. Quadratic model, pure quadratic model, interactions model, linear model and polynomial models are developed for maize yield prediction. A quadratic regression is the procedure of discovering the mathematical equation of the parabola that fits best for a set of information. The pure quadratic regression model mainly has square terms of each attributes. In interactions regression model the dependent variable relies on the interactions between the independent variables. In linear regression the prediction of dependent attribute is done using single independent attributes and not collectively. In polynomial model, the independent attributes are varied up to nthdegree polynomial to predict the outcome. Polynomial models are built by varying their degrees.

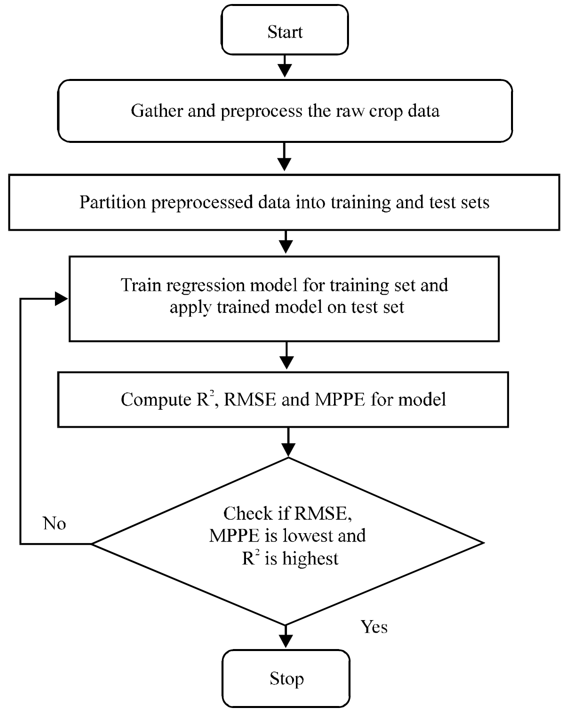

Fig. 1: Regression methodology for crop yield prediction

Table 2 shows the different regression models with the corresponding equations for maize yield prediction. Here x_1 is esw, x_2 is biomass, x_3 is rainfall, x_4 is radiation and y is maize yield. Among all these models, the pure quadratic model was able to accurately predict the maize yield.

The p-values for x_1-x_4 were measured for each the models. It was found that x_2 had the lowest p-value. From this, it can be inferred that biomass (x_2) is contributing more to the maize yield than the other attributes.

Regression model for wheat yield prediction: Here, the wheat data set consists of a total of 50 records. Out of which 35 records are considered as training set and remaining 15 records as test set. Quadratic model, pure quadratic model, interactions model, linear model, polynomial model and Generalized Linear Regression Model (GLM) are built.

Table 3 shows the regression models with the corresponding equations for wheat yield prediction. Here x_1 is biomass, x_2 is Grain-Protein, x_3 is Grain-Size, x_4 is ESW and y is wheat yield.

Table 2: Regression models for maize yield prediction

Models	Equations
Quadratic	$y = 224.87+2.15x_1-0.387x_2-48.04x_3-51.56x_4-0.000085 x_1*x_2-0.32x_1*x_3-0.081x_1*x_4+0.0052x_2*x_3-0.0061x_2*x_4+2.79x_3*x_4+0.00017x_1^2+0.000071x_2^2+3.97x_3^2+1.77x_4^2$
Pure quadratic	$y = -20.8+6.07x_1-0.38x_2-20.92x_3-73.98x_4-0.0184x_1^2+0.000068x_2^2+1.42x_3^2+2.46x_4^2$
Interactions	$y = -774.7+1.29x_1+0.42x_2+36.3x8_3+32.4x_4-0.0007x_1*x_2-0.065x_1*x_3-0.02x_1*x_4-0.0005x_2*x_3-0.012x_2*x_4-0.94 x_3*x_4$
Linear	$y = -26.97-2.45x_1+0.21x_2+23.5x_3+19.9x_4$
Polynomial	$y = 68.2+0.96x_1-0.32x_2-24.68x_3-31.68x_4-0.00021 x_1*x_2+0.000069x_2^2+0.042x_1*x_3+*x_4-0.0066x_2*x_4+0.55x_3*x_4$

Table 3: Regression models for wheat yield prediction

Models	Equations
Quadratic	$y = -864478.68-0.59x_1+100150.25x_2-418784.90x_3+377.801x_4+0.055x_1x_2+8.01x_1x_3-0.00085x_1x_4+29607.81x_2x_3-21.89x_2x_4-222.8x_3x_4+1.131x_1^2-2908.45x_2^2-530884x_3^2-0.00393x_4^2$
Pure quadratic	$y = -845823.73+0.267x_1+101687.6x_2-10573.93x_3+4.46 x_4+0.0000009x_1^2-3060.37x_2^2+915240.24x_3^2-0.0075x_4^2$
Interactions	$y = -94280.1+1.29x_1+5446.13x_2-1205863.77x_3+509.85 x_4-0.054x_1x_2+6.63x_1x_3-0.00089x_1x_4+77463.08x_2x_3-29.8x_2x_4-320.42x_3x_4$
Linear	$y = -9940.39+0.39x_1+474.69x_2+45445.54x_3+1.6x_4$
polynomial	$y =-855791.29-0.25x_1+99469.91x_2-482232.99x_3+347.57 x_4+0.035x_1x_2-2898.28x_2^2+8.36x_1x_3+33562.67x_2x_3-629480.74x_3^2-0.00084x_1x_4-20.19x_2x_4-213.54 x_3x_4$
Generalized Linear Regression (GLM) Model	$y = -927234.86+0.29x_1+111781.56x_2+48076.32x_3+6.16x_4+7.92x_1x_3-0.00068x_1x_4-3376.71x_2^2-1015463.54x_3^2$

GLM models are used when response variables have arbitrary distributions (rather than simply normal distributions). Since, the response variable here is the wheat yield which can vary arbitrarily this model is more suited than the other regression models. After measuring the p-values it was found that biomass was contributing more to the wheat yield. Also GLM performed better than the other models.

Regression model for cotton yield prediction: Here, the cotton data set consists of a total of 123 records. Out of which 84 records are considered as training set and remaining 39 records as test set. Quadratic model, pure quadratic model, interactions model, linear model, polynomial model and Stepwise Linear Regression Model (SLM) are built.

Table 4 shows the regression models with the corresponding equations for cotton yield prediction. Here x_1 is dw_total, x_2 is NUptake, x_3 is bolls_sc, x_4 is lai_max, x_5 is esw, x_6 is rainfall, x_7 is radiation, x_8 is maximum temperature, x_9 is minimum temperature and y is the cotton yield.

In SLM, independent variables are included in the regression model step by step. The variables which contribute significantly to the outcome are retained in the model. Another method is, to include all independent variables in the model. Later variables which are not significant are eliminated from the model. After measuring the p-values it was found that dw_total was contributing more to the wheat yield. Also SLM performed better than the other models.

Table 4: Regression models for cotton yield prediction

Models	Equations
Quadratic	$y = 11674.19+3.26x_1- 118.65x_2+917.16x_3-1001x_4+29.67x_5-62351.39x_6+66.18x_7+68.89x_8+150.77x_9+0.014 x_1x_2-0.38x_1x_3+0.71x_1x_4-0.0041x_1x_5+0.05x_1x_6+0.021x_1x_7+0.074x_1x_8-0.097x_1x_9+19.77x_2x_3-41.37x_2x_4+0.18 x_2x_5-35.99x_2x_6+0.17x_2x_7-5.19x_2x_8+5.16x_2x_9-258.38 x_3x_4-1.79x_3x_5+10544.32x_3x_6-10.16x_3x_7-2.16 x_3x_8-1.44 x_3x_9+1.59x_4x_5-7617.05x_4x_6-104.1x_4x_7+159.16 x_4 x_8-68.85 x_4x_9-9.62x_5x_6-0.056x_5x_7-0.072x_5x_8-0.10x_5x_9-546.67x_6x_7+2253.38x_6x_8-2713.98x_6x_9-4.65x7x_8+6.71 x_7x_9-8.43x8x_9-0.00025x_1^2-0.12x_2^2+25.27x_3^2+306.47x_4^2 -0.0204x_5^2+1022.49x_6^2-1.17x_7^2+4.54x_8^2+0.105x_9^2$
Pure quadratic	$y = -747.22+1.06x_1-24.83x_2-579.39x_3-301.39x_4+4.79x_5-18.93x_6+17.09x_7-37.83x_8+62.29x_9-0.0000449x_1^2+0.041x_2^2+83.11x_3^2+56.17x_4^2-0.0057x_5^2+ 0.43x_6^2-0.75x_7^2+0.88 x_8^2-1.96x_9^2$
Interactions	$y = -9724.84+2.77x_1-121.66x_2+1380.23x_3+1120.51x_4+9.14x_5-510.44x_6+226.03x_7+179.002x_8-71.15x_9-0.00086 x_1x_2-0.43x_1x_3-0.47x_1x_4+0.00072x_1x_5+0.059 x_1x_6-0.039 x_1x_7+0.042x_1x_8-0.015x_1x_9+23x_2x_3+19.36x_2x_4-0.0088 x_2x_5-4.15x_2x_6+1.88x_2x_7-2.35x_2x_8+0.82x_2x_9-217.3x_3x_4-1.83x_3x_5+69.25x_3x_6+3.31x_3x_7-34.26x_3x_8+23.46x_3x_9-1.82x_4x_5+245.62x_4x_6-14.98x_4x_7+50.21x_4x_8-18x_4x_9+0.16x_5x_6-0.37x_5x_7+0.23x_5x_8-0.102 x_5x_9+6.76x_6x_7-4.87x_6x_8+6.7x_6x_9-5.15x_7x_8+1.97x_7 x_9-0.25x_8x_9$
Linear	$y = -828.88+0.23x_1-2.31x_2+274.81x_3-104.007x_4-1.005 x_5-5.13x_6-0.411x_7-3.03x_8+29.69x_9$
Polynomial	$y = -10903.21+4.22\times1-173.6x2+1384.44x3+96.62\times4+9.67\times5-356.98\times6+167.36\times7+257.54\times8-68.25\times9-0.000121\times1^2+0.0047\times1\times2-0.5\times1\times3+25.82\times2\times3-0.1671\times4+3.84\times2\times4-194.06\times3\times4-0.0002\times1\times5+0.023\times 2\times5-1.69\times3\times5-0.802\times4\times5+0.047\times1\times6-2.93 \times2\times6+38.49\times3\times6+170.74\times4\times6+0.13\times5\times6-0.05\times1\times7+2.42 \times2\times7+7.91\times3\times7-23.51\times4\times7-0.29\times5\times7+7.77\times6\times7+0.05\times1\times8-3.13\times2\times8-38.12\times3\times8+94.02\times4\times8+0.14\times5x8-6.91\times6\times8-5.31\times7\times8-0.038\times1\times9+2.006\times2\times9+24.01\times3\times9-44.98\times4x9-0.09\times5x9+7.96\times6\times9+2.96\times7\times9-1.15\times8\times9$
Stepwise Linear Regression	$y = -251.16+1.54x_1-50.61x_2-561.84x_3+65.07x_4+0.14x_5-6.29x_6-7.61\times7+9.27\times8+47.74\times9-0.0075\times1\times2- 3.16\times2\times4+0.32\times2^2+81.02\times3^2+67.01\times4^2-1.28\times9^2$

RESULTS AND DISCUSSION

Experimental setup and results: This study explains the experiments carried out on wheat, cotton and maize data sets using quadratic, pure quadratic, linear, polynomial, generalized linear regression and stepwise linear regression models. It also compares the results obtained from them. Implementation is done on Windows 7 operating system using MATLAB R2013a as the programming tool.

Accuracy of these prediction models are measured using (R^2), Root Mean Square Error (RMSE) and Mean Percentage Prediction Error (MPPE).

The R^2 is a statistical measure of how close the data are to the fitted regression line. It is given by Eq. 1 (Nancy, 2010):

$$R^2 = 1 - (n - 1/n - p)(SSE/SST) \quad (1)$$

Where:

SSE = Sum of Squared Error
SST = Sum of Squared Total
n = The number of observations
p = The number of regression coefficients (including the intercept)

If the value of R^2 is closer to one then it is able to explain most of the variations on the prediction model (Nancy, 2010). RMSE provides the difference between predicted and actual values. It is given by Eq. 2:

$$RMSE = \sqrt{\sum_{t=1}^{n}(y_t - y)^2 / n} \quad (2)$$

Where:

y_t = The predicted value
y = The actual value
n = The number of samples

MPPE provides the average of percentage prediction errors. It is given by Eq. 3:

$$MPPE = 1/n \sum_{t=1}^{n}(y - y_t)/y \times 100 \quad (3)$$

Where:

y_t = The predicted value
y = The actual value
n = The number of samples

Prediction results for wheat yield prediction: Figure 2 shows the results based on R^2 statistic and MPPE values for wheat yield prediction using linear, pure quadratic, interactions, quadratic, polynomial and GLM models. As can be observed, the R^2-value for the GLM regression model is higher than the rest of the models while MPPE is lower. From this it can be inferred that the GLM model accurately predicts the wheat yield than the other models.

Figure 3 shows the results of RMSE value for wheat yield prediction. It clearly shows that the GLM model has a lower RMSE value than the other models.

Prediction results for maize yield prediction: Figure 4 shows the results based on R^2 statistic and MPPE, for maize yield prediction using linear, pure quadratic, interactions and quadratic and polynomial models. As can be observed, the R^2 for the pure quadratic regression model is higher while MPPE is lower than the rest of the models. From this it can be inferred that the pure quadratic model accurately predicts the wheat yield better than the other models.

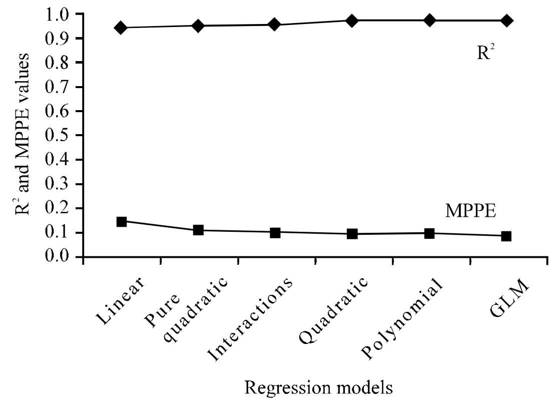

Fig. 2: Accuracy graph for wheat yield prediction based on R^2 (MPPE)

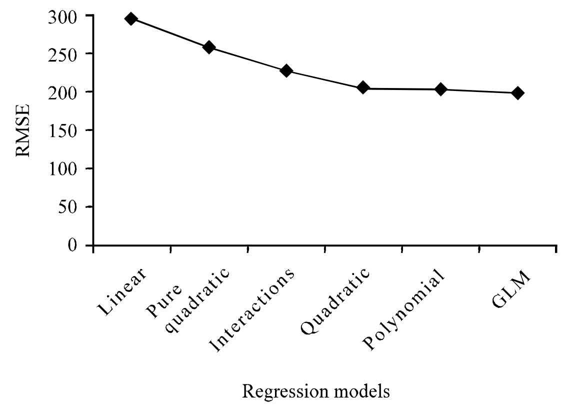

Fig. 3: Accuracy graph for wheat yield prediction based on RMSE

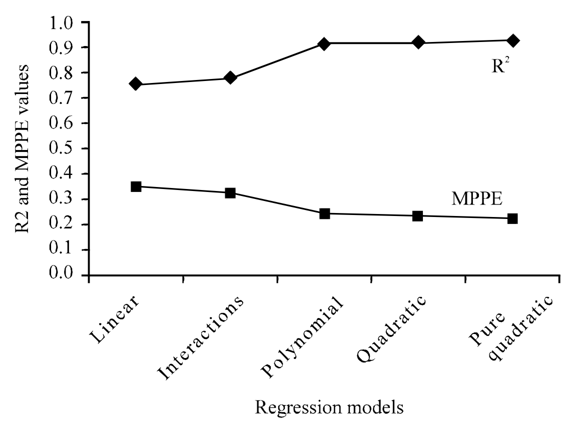

Fig. 4: Accuracy graph for maize yield prediction based on R^2 (MPPE)

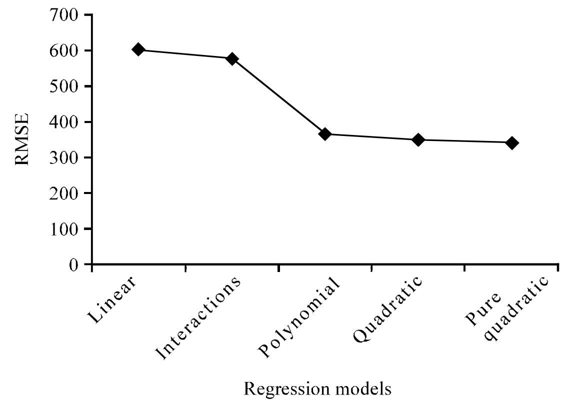

Fig. 5: Accuracy graph for maize yield prediction based on RMSE

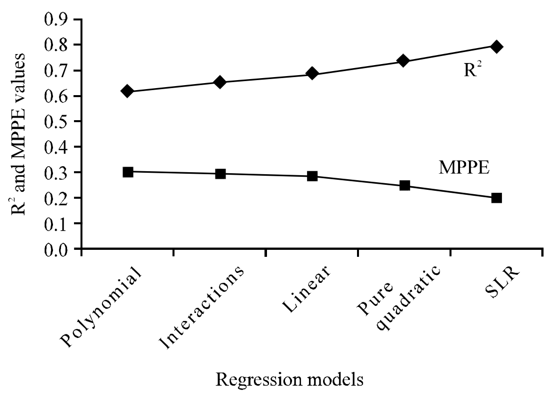

Fig. 6: Accuracy graph for cotton yield prediction based on R^2 and MPPE

Figure 5 shows the results of RMSE value for maize yield prediction. It clearly shows that the pure quadratic model has a lower RMSE value than the other models

Prediction results for cotton yield prediction: Figure 6 shows the results based on R^2 Statistic and MPPE, for cotton yield prediction using linear, pure quadratic, interactions, quadratic, polynomial and proposed SLR regression models. As can be observed, the R^2-value for the SLR model is higher while MPPE is lower than the rest of the models. From this it can be inferred that the SLR model accurately predicts the cotton yield better than the other models.

Figure 7 shows the results of RMSE value for cotton yield prediction. It clearly shows that the proposed SLR regression model has a lower RMSE value than the other models.

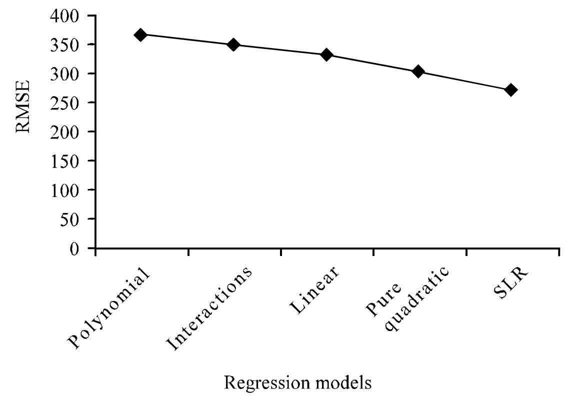

Fig. 7: Accuracy graph for cotton yield prediction based on RMSE

CONCLUSION

The effort demonstrated that regression techniques can be utilized for yield prediction for the area with satisfactory results. India is a standout amongst the most of the nations in creating yields in Asia and utilization of wheat, Maize, cotton in numerous piece of our nation is seen broadly. Thus, an effort made to predict the yield of such in India. In order to predict the yield, Regression model is utilized as a prediction tool and some of the important factors in yield production are selected. The information gathered along with the different attributes is used as the input variables for the regression model. Along these lines, the best regression model is found for the yield. Every model is run a few times to deal with conceivable estimations of root mean square and R^2 statistics values. By utilizing the best regression model for the survey, the forecast of generation of wheat, maize and cotton is done for chosen years. The outcomes demonstrate that the proposed regression model is a suitable method for foreseeing yield production. The results of different models are compared based upon the root mean square, R^2 statistics and percentage prediction error. The model which gives the lower Root mean square, percentage prediction error and Higher R^2 statistics values is considered to be the best model for crop yield prediction.

RECOMMENDATIONS

In future, this research can be extended by applying different prediction techniques like Support Vector Regression (SVR), Neural Networks, Fuzzy Logic, etc. for predicting the yield of various crops. Further, co-relation between predictor variables may be found which gives the importance of the variable for prediction.

REFERENCES

Alan, O.S., 1993. An introduction to regression analysis. Master Thesis, University of Chicago Law School, Chicago, Illinois.

Asseng, S., G.C. Anderson, F.X. Dunin, I.R.P. Fillery and P.J. Dolling *et al*., 1997. Use of the APSIM wheat model to predict yield, drainage and NO3-leaching for a deep sand. Crop Pasture Sci., 49: 363-378.

Carberry, P.S. and D.G. Abrecht, 1991. Tailoring Crop Models to the Semi-Arid Tropics. In: Climatic Risk in Crop Production: Models and Management in the Semi-Arid Tropics and Sub-Tropics, Muchow, R.C. and J.A. Bellamy (Eds.). Cab International, Wallingford, England, pp: 157-182.

Hearn, A.B., 1994. OZCOT: A simulation model for cotton crop management. Agric. Syst., 44: 257-299.

Nancy, R.Z., 2010. Topic: Multiple Linear Regression. Stanford University, California.

Qaddoum, K. and E.L. Hines, 2012. Reliable yield prediction with regression neural networks. Proceedings of the 12th WSEAS International Conference on Signal Processing, Computational Geometry and Artificial Intelligence, August 21-23, 2012, WSEAS Press, Turkey, Istanbul,-pp: 1.

Sanchez, A.G., J.F. Solis and W.O. Bustamante, 2014. Attribute selection impact on linear and nonlinear regression models for crop yield prediction. Sci. World J., 2014: 1-10.

Zaefizadeh, M., A. Jalili, M. Khayatnezhad, R. Gholamin and T. Mokhtari, 2011. Comparison of Multiple Linear Regressions (MLR) and Artificial Neural Network (ANN) in predicting the yield using its components in the hulless barley. Adv. Environ. Biol., 10: 109-114.

Zaw, W.T. and T.T. Naing, 2009. Modeling of rainfall prediction over Myanmar using polynomial regression. Proceedings of the International Conference on Computer Engineering and Technology, January 22-24, 2009, IEEE, New York, USA., ISBN: 978-0-7695-3521-0, pp: 316-320.

Zhang, L., L. Lei and D. Yan, 2010. Comparison of two regression models for predicting crop yield. Proceedings of the IEEE International Symposium on Geoscience and Remote Sensing (IGARSS), July 25-30, 2010, IEEE, New York, USA., ISBN: 978-1-4244-9565-8, pp: 1521-1524.