IARJSET

ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Vol. 8, Issue 8, August 2021

DOI: 10.17148/IARJSET.2021.8855

Analyzing Agricultural Crop Production and their Uncertainty Using Linear Regression and Fuzzy Logic

Om Prakash Singh[1], Bijay Kumar Mandal[2], Sunil Kumar[3]

[1]Assistant Professor –Department of Computer Science and Engineering, Vidya Vihar Institute of Technology, Purnea, Bihar

[2]Assistant Professor – Department of Mathematics, Vidya Vihar Institute of Technology, Purnea, Bihar

[3]Research Scholar – Computer Applications, Manipal University, Jaipur, Rajasthan

ABSTRACT

Our nation India is an agrarian country and its economy based upon agricultural crops production. The share of agriculture in GDP increased to 19.9 percent in 2020-21 and almost fifty percent of total manpower utilized their efforts in this sector. The uncertain whether, climate changes and water storage with traditional farming trends and improper irrigation facilities are directly affect the crop productivity. All such parameters make the environment of uncertainty regarding crops production. On the other hand, accurate and timely predictions of crop production are backbone of the policy maker regarding import-export, demand-supply, marketing, pricing and distributions to balance the socio-economic frame.

The uncertainty and its prediction tend to be complex phenomena. The primary resources of uncertainty are randomness and fuzziness. The randomness deals with general uncertainties while fuzzy logic suitable for the complex phenomena. The statistical regression methodology is used traditionally for such complex predictions. In series of smart notations, the smart forming is necessity of day. Hence there is a requirement to develop qualitative and statistically sound prediction of crop yields with machine learning handling large amount of data.

The present works focus on investigation of various used machine learning algorithms for their suitability in crop yields perdition and finally proposed an approach based on linear regression and fuzzy logic in big data computing paradigm for accurate and timely predictions of crop production.

Keywords: Uncertainty, Computational Intelligence, Linear Regression, Fuzzy Logic, Time Series Analysis

1. INTRODUCTION

The factors like climate, weather, geographical, topography, historical, biological, institutional, political and socio-economic are affect the Indian agriculture for both food crops and commercial crops. As visible, these factors are independent from each other. Due to change in natural factors and technology with time, the policies also reforms. All such variations root cause of change in agriculture crop production performance. Therefore the crises in consistent output of food and corresponding risk arise. The production is mostly affected by environmental factors because weather influences the crop growth and development. The other factors like soil properties, fertilizer, irrigation, tillage etc affect the productivity. There is a non linear relationship between crop yield and these factors. The uncertainty is associated with such factors and as well as crop yield.

.

Early accurate prediction of crop production prior to harvest is an important issue in agriculture because it would facilitate the formers to make the precautionary actions regarding selection of crops and improving the productivity with better vision on cultivation of seasonal crop and its scheduling and changes in crops yields influence the foods demand- supply, marketing, pricing and distribution in local and global market both. It also helps to policy makers. Such prediction is possible with collection of previous experiences of the formers and others influencing factors stored as a large database say bigdata. The common input parameters are farm capacity, soil feature, rainfall, humidity, temperature, solar radiation, irrigation, fertilizer used and tillage. The machine learning techniques might be efficient for crop yield early prediction. The most common techniques are classification, regression and clustering. The intelligence computing and statistics brought by machine learning to improve the prediction power. But our focus terminologies are regression and fuzzy logic to resolve such a critical problem.

[Prof. L. Zadeh, 1965] proposed the fuzzy logic as an extension of binary logic founded on fuzzy set theory. It is a many value logic that deals with situations where boundaries are not well determined. In such cases, the linguistic labeling is used for mathematical purpose. It helps in expressing an uncertainty about tangible meaning of used labels and tolerates soft constraints and flexible requirements. It is also the tool for reasoning. **[C D Cox, 1992]** The fuzzy set theory, Artificial intelligence, Neural Network and Genetic algorithm come closer and make a platform called computational intelligence. The regression extracts valuable information from bigdata, predicts and modeled the fuzziness.

2. AGRICULTURE AND MACHINE LEARNING

Machine learning is a method of data analysis that automates analytical building. It is a branch of artificial intelligence based on concept that a model can learn from data, identify patterns and make decisions with minimal human intervention. The machine learning tasks are classification, regression and clustering. Early prediction of crop production can be performed by using various machines learning algorithms based on machine learning tasks like classification, regression and clustering. The some past work related to it are considering for description in support of our consideration for linear regression.

2.1. Classification task

The classification task deals with the use of machine learning algorithm that learns how to assign a class label to examples from the problem domain. There are many different types of classification tasks such as binary, imbalanced, multi class and label classification. There are a number of series of various machine learning algorithms used for classification purpose. The most commonly used algorithms for the purpose of prediction of crop production are describing in subsequent paragraphs.

[C Chen & H Mcnairn, 2006] develop a rice crop monitoring system based on Artificial Neural Network (ANN). The neural networks are better in intelligence computing over the conventional methods specially problem related to prediction. This monitoring system deals with neural network classification. The system was used in rice production areas in both wet and dry season for time variant as different planting dates than it was able to extract information with minimum accuracy mapping of 96%. The predicted yields and government statistics value are almost the same. **[K P Brindha, 2013]** adopted principal components analysis based data mining process for correct forecasting about time based development of monsoon rainfall in India while the statistical method has issue of clear cut applications. **[Y. Yiqun, W James & M McNicol, 1994]** were used the Bayesian belief network for study of effect of climate change on potato production and finding was related to uncertainty of future climate change with temperature, rainfall, sun radiation and knowledge about potato development as input parameters. This outcome is compared with conventional mathematical method output that votes the performance of Bayesian network.

2.2. Regression task

The regression task is a process of finding the correlation between dependent and independent variables. It helps in predicting the continuous variables so the task of regression based machine learning algorithms are finding the mapping function to map the input variable to the continuous output variable. There are also various regression based algorithms but our consideration is linear regression which is discussed in section 3.

[M Manish Kaul, L Robert & H Hill, C Walthall, 2005] uses the ANN and multiple linear regressions for developing a model for corn and soybean crops yield forecasting with climate aspect. They show that ANN model gives more accurate yield prediction than multiple linear regressions. Although this favoring to ANN but for a multiple parametric application, it setup a mile stone for use of linear regression. **[P R Prasad & S A Begum, 2013]** uses linear regression and feed forward neural networks models for prediction of crop production. They conclude that ANN is more suitable when relationship between the variables is unknown and complex, and very difficult to handle statistically. But the linear regression can be used when the variable are known.

2.3. Clustering task

The clustering task related to identifying the objects that are similar to each other but different from individuals in other

DOI: 10.17148/IARJSET.2021.8855

groups. It basically divides the data in to number of groups based on some similarity and dissimilarity. The K-Means and K- Medoids are most commonly used machine learning clustering algorithm.

[P Utkarsh, N Narkhede, K P Adhiya, 2014] demonstrates the crop prediction based on K- Means clustering algorithm having maximum number of high quality clusters, correct prediction and maximum accuracy outcome. On the other hand the study from **[R Glauston, T D Liman & S Stephany, 2013]** demonstrates about classification of metrological data. For this purpose, they are used the frequency of variable from weather forecast model.

3. LINEAR REGRESSION AND FUZZY LOGIC

As we know the regression is one of the most widely used statistical methods for representation of relationship among variables and predict uncertain phenomena. It is well suited for model determination in a situation where observed and estimated value is different due to measurement error and random variations. But this difference is due to imprecise observed data or parameters and structure of the system.

[A G Sanchez, J F Solis & W O Bustamante, 2014] suggested the linear regression due to its usability, it is a statistical technique applied on linear system that capable to measure the relationship between dependent and independent variables. It is used as multiple regressions in the case of two or more input attributes associated with an independent variable and show the consistent result in standards test. It has a limitation with regression assumption for multiple co linearity among dependent and independent variables. But the two study **[D Ramesh & B V Vardhan, 2015]** and **[J K Betty G Shem & Sitienei, 2017]** make it important. First is associated with crop yield prediction using data mining and second is with use of regression models for prediction of tea crop yield.

The fuzzy logic is a branch of machine intelligence that allows expressing and evaluating complex relation in a linguistic way in an uncertain world. As discussed earlier, it is an efficient technique to put knowledge right in to a technical solution and intimates the human observation, action and decision process in a system. Mathematically it is from logic section that uses the degree of membership in sets rather than a strict true and false membership or crisp value. The fuzzy logic utilized its capability to use diverse aspect of vagueness of daily life activities, real time problems, control system and many more as expert. The fuzzy logic, regression technique and its allied aspect of prediction play an important role by extracting valuable information from imprecise data, predict and model the fuzzy objects.

We have basic task to utilized joint capability of both linear regression and fuzzy logic to handle the big data for crop prediction purpose. **[J V Tu, 1996; Dreiseitl & O Machado, 2003; M Nasiri, 2005; F Shapiro, 2005]** discussed about the existing prediction techniques based on statistical and artificial intelligence such as ANN, support vector machine, K- nearest neighbors, Fuzzy logic, Fuzzy Neural Network and Fuzzy regression for the prediction of uncertainties with their drawbacks and limitations. The statistical techniques are not suitable for inadequate number of observation and distribution assumptions. On the other hand, artificial intelligence prediction techniques has limitations as low interpretation ability, less and slow identifying the relationship between variables, over fitting, lack in flexibility to incorporate new knowledge and unsuitability with high dimensional data.

[H. Tanaka, S Uejima & K Asai, 1982] were introduced the Fuzzy linear regression with its virility as estimating the relationship between variables in a very limited and imprecise data environment, variables are interrelated with uncertainty and relationship between the variable can be interpreted qualitatively. The regression is a mathematical technique used for model the relationship between explanatory and response variables. **[F Shapiro, 2005; R M Dom, 2007, 2008; J Miles & M Shevlin, 2001& A Bisserier, 2010]** are agree with two types of fuzzy regression approaches one as Tanaka's Linear Programming Approach and another is Fuzzy Least Square Approach. The fuzzy linear regression uses the prediction interval in machine learning and linear function only.

[A Bisserier, 2010] urged that fuzzy regression more useable in evaluation of functional relationship between dependent and independent variables in a fuzzy environment so it offers an efficient tools for analyzing complex phenomena having vague and imprecise quantitative and qualitative data. The fuzzy regression is none statistical method that based on possibility and fuzzy set theory while statistical regression based on only probability theory.

4. ANALYZING METHODS

The analysis is carried out with data set, since crop predictions depends upon numbers of diverse factors so we have the big data having structured and unstructured both data. The other parameters for such analysis are fuzzy linear regression analysis and time series analysis that help is projection of proposed approach.

DOI: 10.17148/IARJSET.2021.8855

4.1. Data Collection

It is hard to just collect data and use it this is because we have target to time series analysis also. It is always best to use the season by season, year by year, and decade by decade data are used formulate the training of machine learning algorithm for accurate and timely prediction of crops yield. The different government and non government keep the data. Few of them are agriculture production data from spriter-GIS, weather information data from SMN, CRU TS 3.24.01, State government records, publicly available Indian government records, indiawaterport1.org etc. The most commonly used tools are MATLAB, Python, R, SPSS and WEKA with different methods as described in section 2 are just sufficient to dry run of proposed approach.

4.2. Fuzzy Linear Regression Analysis

The fuzzy linear regression can be viewed as two categories. The first category possibilistic regression analysis (PRA) and second one is fuzzy least square method (FLSM). The fuzzy logistic regression is based on PRA while fuzzy linear regression is on FLSM.

[Tanaka, 1989] described that the possibilistic regression analysis is based on possibility theory. It uses the fuzzy linear system as a regression model for minimizing the total uncertainty of the estimated values of dependent variables. Since the membership function of fuzzy sets is described as interactive possibility distribution and fuzzy coefficients are determined independently of each other. Therefore the interactive possibility distribution of coefficient of the fuzzy linear system is replace by quadratic membership functions and exponential possibility distributions to formulate the possibilistic regression analysis. The other thing that used in formulating the possibilistic regression is inclusion relation between the given outputs and the estimated outputs. This is applicable to linear function only. If the real valued input vector of independent variables are X= [X0, X1, X2, X3,….Xk]^T and regression coefficients are Ai, where i= 0, 1 ,2 …k as the symmetric triangular fuzzy number with center Ci and half width Wi, than the fuzzy output Y is written as

Y= A0.X0+ A1.X1+A2.X2+……..Ai.Xi+…Ak.Xk

[H. Tanaka, S Uejima & K Asai, 1982] proposed the fuzzy linear regression that adopts the FLSM. It minimizes error between estimated and given outputs. It has minimum degree of fuzziness between observed and estimated values while PRA has simplicity in programming and computation. The purpose of such regression is describing the variation in dependent variable in terms of variations in independent variable. If dependent variable is y and dependent variable is x than y= f(x). Here f(x) is a linear function.

4.3. Time Series Analysis

The ability to predict about the future is known as extrapolation in the classical statistical handling of time series data. It is a method to analyze time on parametric and series data to extract the characteristics of the data and predict future values based on previously observed values. It is an important tool for crop production prediction in which crop yield is a dependent variable that considered as a time function. This consideration provides the relationship between crop yield and time. The time series analysis consists of various methods like linear and nonlinear, parametric and non-parametric, univariate or multivariate and time domain or frequencies domain**[L Hong Ying, H Yan Lin, Y Yong Juan and Z Hui Ming, 2012].**

4.4. Prediction Accuracy Validation

As we have predicted value against the some observed value in crop yield prediction process. Now it is necessary to evaluate the accuracy of this prediction. This evaluation is actually the prediction accuracy validation. The accuracy of prediction can only be determined on the basis of prediction model performance with new data. We can carry out this validation with both training and test data approach and scale-dependent error approach.

With training and test data approach, the available data are put in to partition one as training data and another as test data. Almost the size of test data set is 20% of total data in common practice. The training data is used to estimate the used parameters while test data is used to evaluate the accuracy of prediction method.

The scale-dependent error approach uses the root mean squared error (RMSE), mean squared error (MSE) and mean absolute error (MAE). The RMSE is the standard deviation of the prediction error in which the difference of predicted values and observed values is squared and averaged and then mean value is computed. The MSDE compute the square of difference of predicted values and observed values and then average those values. Finally the MAE is computed as the average of absolute difference of predicted values and observed values.

5. PROPOSED APPROACH

In this article the machine learning based linear regression with fuzzy logic for accurate and timely prediction of crops production is discussed from model selection and data acquisition to its validation. The proposed approach framework consists of fuzzy logic based controller- Fig1, Fuzzy Linear regression-Fig2 and Framework-Fig3. The factors or parameters that influence the crop yields are uncertain by nature so the probability will be computed by using linear regression. Then fuzzy logic will be applied for various parameters.

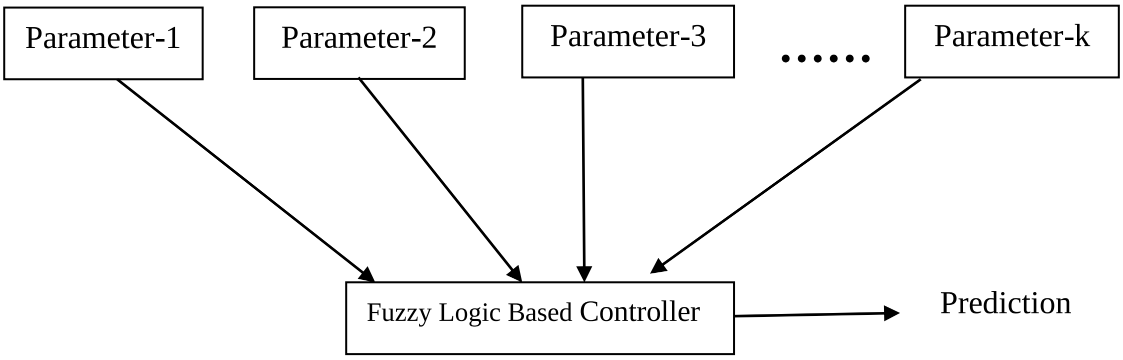

Fig-1

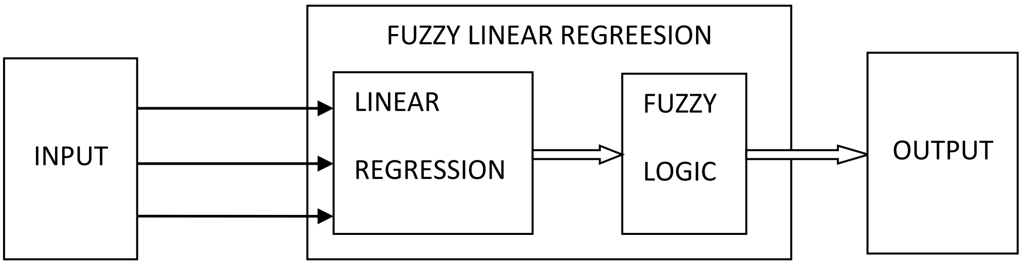

Fig-2

As our consideration the data as big data, therefore the process of prediction will be use this data at two level, first in preprocessing and second at prediction.

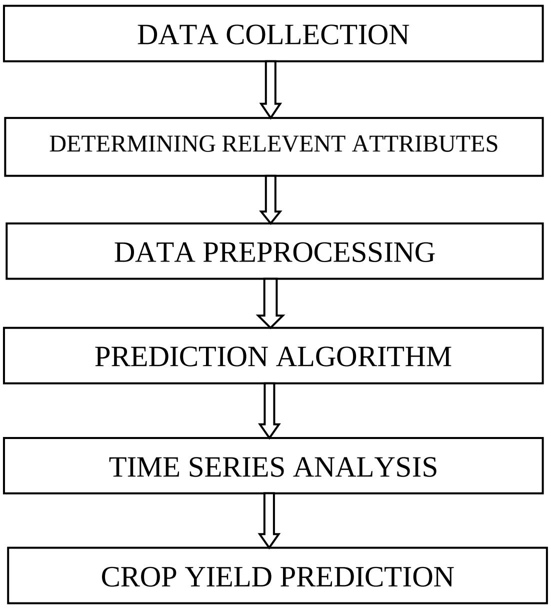

Fig-3

DOI: 10.17148/IARJSET.2021.8855

6. CONCLUSION

There is necessity of employed the machine learning techniques in agriculture for finding the many secret knowledge by analyzing such big data where data are structured and unstructured with insufficient relationship between dataset. The various literatures show that now the different machine learning techniques used for predictions of crop yield but this proposed approach capable for accurate and timely predictions of crop yield production, empowering the former community and balancing the socio-economic frame. The fuzzy logic and regression derived linear fuzzy regression is most suitable alternative of statistical regression model to handle the diverse qualitative and quantitative parameters as a real time application. We are in our best effort to implement this approach with Python.

7. REFERENCES

[01] H Tanaka, S Uejima, & K Asai, "*Linear Regression Analysis with Fuzzy Model*", IEEE Transactions on Systems, Man and Cybernetics, Vol. 12, No 6, 1982, pp 903 – 907.

[02] E D Cox, "*Integrating Fuzzy Logic with Neural Networks"*, AI Expert, 1992, pp.40-45.

[03] Y Yiqun Gu, W James & M McNicol, *"An Application of Belief Networks to Future Crop Production"*, IEEE Conference on Artificial Intelligence for Applications, San Antonia, TX. 1994. p. 305–9.

[04] J V Tu, *"Advantages and Disadvantages of Using Artificial Neural Networks Versus Logistic Regression for Predicting Medical Outcomes"*, J Clin Epidemol, Vol 49, No 11, 1996, pp.1225-1231.

[05] O Yun-His, Chang, M Bilal & Ayyub, *"Fuzzy regression methods- a comparative assessment, Fuzzy Sets and Systems"* 119, 2001, pp. 187-203.

[06] J Miles & M Shevlin, *"Applying Regression and Correlation. A guide for Students and Researchers"*, SAGE Publication Ltd., 2001.

[07] S Dreiseitl & O Machado, *"Logistic Regression and Artificial Neural Network Classification Models: a Methodology Review"*, Journal of Biomedical Informatics, 35, 2003, pp352- 359.

[08] M Nasiri *"Comparison of Statistical Regression, Fuzzy Regression and Artificial Neural Network Modeling Methodologies in Polyester Dyeing"*, Proceedings of 2005 International Conference for modeling, control and automation, 2005.

[09] M Monisha Kaul, L Robert, H Hill & C Walthall, *"Artificial neural networks for corn and Soybean yield prediction"*, Elsevier. Agricultural System. 2005; 85(1):1–18.

[10] M Nasiri, *"Comparison of Statistical Regression, Fuzzy Regression and Artificial Neural Network Modeling Methodologies in Polyester Dyeing",* Proceedings of 2005 International Conference for modeling, control and automation, 2005.

[11] C Chen & H Mcnairn, *"A neural network integrated approach for rice crop monitoring. International Journal of Remote Sensing",* 2006; 27(7):1367–93.

[12] Rosma M Dom, *"An Adaptive Fuzzy Regression Model for the Prediction of Dichotomous Response Variables",* Fifth International Conference on Computational Science and Applications, IEEE, 2007.

[13] Rosma M Dom, *"A Learning System Prediction Method Using Fuzzy Regression"*, Proceedings of the International Multi Conference of Engineers and Computer Scientists, Vol- I,IMECS- 2008, Hong Kong.

[14] Rosma M Dom, *"A Learning System Prediction Method Using Fuzzy Regression"*, Proceedings of the International Multi Conference of Engineers and Computer Scientists, Vol- I,IMECS- 2008, Hong Kong

[15] Amory Bisserier, *"A revisited approach to linear fuzzy regression using trapezoidal fuzzy Intervals"*, Information Sciences, 180, 2010, pp. 3653–3673

[16] S Brdar, D Culibrk, B Marinkovic, J Crnobarac and V Crnojevic, *"Support Vector Machines with Features Contribution Analysis for Agricultural Yield Prediction"*, in the Proc. of Second International Workshop on Sensing Technologies in Agriculture, Forestry and Environment, 2011.

[17] T D Dongale, T G Kulkarni, & R R Mudholkar, *"Fuzzy Modelling of Voltage Standing Wave Ratio using Fuzzy Regression Method"*, International Journal of Emerging Technology and Advanced Engineering, ISSN 2250-2459, Volume 2, Issue 6, 2012, pp. 21-26.

[18] L Hong-Ying, H Yan-Lin, Y Yong-Juan & Z Hui-Ming, *"Crop yield forecasted model based on time series techniques",* Journal of Northeast Agricultural University (English Edition). 2012; 19(1):73–7.

[19] K P Brindha, *"Data mining based on principal component analysis for rainfall forecasting in India"* . International Journal of Advanced Research in Computer Science and Software Engineering (IJARCSSE). 2013 Sep; 3(9):1–11.

[20] P R Prasad & S A Begum, *"Regression and neural networks models for prediction of crop production"*. International Journal of Scientific and Engineering Research. 2013 Sep; 4(9):98–108.

[21] R Glauston, T D Liman & S Stephany, *"A new classification approach for detecting severe weather patterns"*, Computers and Geosciences, ELSEVIER. 2013; 57:158–65.

[22] A Gonzalez-Sanchez, J Frausto-Solis & WOjeda-Bustamante, *"Attribute selection impact on linear and nonlinear regression models for crop yield prediction",* Sci. World Journal, 2014.

[23] P Utkarsha, N Narkhede &, K PAdhiya, *"Evaluation of Modified K-Means Clustering Algorithm in Crop Prediction"*. International Journal of Advanced Computer Research. 2014; 4(3):1–1.

[24] D Ramesh and B Vishnu Vardhan,"Analysis Of Crop Yield Prediction Using Data Mining Techniques", International Journal of Research in Engineering and Technology, 2015, 4(1)

[25] J Kenya Betty, Sitienei , Shem G Juma, and Everline Opere, *"On the Use of regression Models to Predict Tea Crop Yield Responses to Climate Change: A Case of Nandi East, Sub-County of Nandi County"*, MDPI Sensors, 2017

[26] Y X Su, H Xu & I J Yan, *"Support vector machine-based open crop model (SBOCM): Case of rice production in China*". Saudi J Biol Sci. 2017;24(3):537–547. doi:10.1016/j.sjbs.2017.01.024.

[27] Arun Kumar, Naveen Kumar, Vishal VatArun Kumar, Naveen Kumar & Vishal Vats, *"Efficient Crop Yield Prediction Using Machine Learning Algorithms",* International Research Journal of Engineering and Technology, 05(06), 2018.

[28] Rushika Ghadge, Juilee Kulkarni, Pooja More, Sachee Nene &R L Priya, *"Prediction of Crop Yield using Machine Learning"*, International Research Journal of Engineering and Technology, vol. 5, Issue 2, Feb-2018, pp.2237-223