International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-4, November 2019

Prediction of Sugarcane Yields from Field Records using Regression Modeling

Shivani S. Kale, Preeti S. Patil,

Abstract: Prediction Of Sugarcane Crop Yield Benefits The Farmer To Get Best Possible Decision Regarding Sugarcane Crop Cultivation. The Purpose Of This Work Is To Identify Possible Relationship Between N, P, K Fertilizer, Water Resource And Planting Densities.
The Algorithm Used Is Multiple Regression. The Paper Focuses On The Generation Of Multiple Regression Models For The Dataset Of Sugarcane Crop For Season Adasali, Suru And Preseasonal Method. The Intercept And Slope For Variables Are Calculated And Equation For Each Model Is Generated. Sample Of N,P,K And Other Are Considered For A Period Of 7 Years From 2012 To 2018. Data Of Experimentation Is Collected For Arid Region I.E. Pandharpur, Maharashtra State.

Keywords: Prediction, Crop Yield, Regression Analysis

I. INTRODUCTION

As we know India is the land of agriculture, approximately 70 % of population constitutes farmers. As in last few years the extensive boost of population raises the question of fulfillment of food requirement. So the slope of production needs to be at increasing order. Due to technological enhancement in agriculture practices which results in increased in production for initial certain years. But after some years usage of fertilizer and unpredictable weather conditions has make production of crop as challenging issue. So the improvement in crop production to contribute to national income is needed. So prediction of yield and recommendation of fertilizer will surely help the farmer for getting good output yield. The multiple regression algorithm considers parameters such as N, P, K to predict yield and try to find out the relationship with the input parameters and output parameters. The data is visualized with respect to different parameters using juypter, python. As it will help to make clear parameters dependencies with output and other variable. There are four methods of Sugarcane cultivation in Maharashtra. Maharashtra is the second largest producer of sugarcane in India. The research for increasing sugarcane production will be helpful to gain more production.

Manuscript published on November 30, 2019.
* Correspondence Author
Shivani S. Kale*, Assistant professor, KIT, Kolhapur, Maharashtra, India
Dr. Preeti S. Patil, Assistant professor, KIT, Kolhapur, Maharashtra, India

The fertilizer recommendation for sugarcane crop can be done by considering N, P, K parameters as well as Ph value of soil. This will surely help in cost cutting of fertilizer usage and more production as well as land fertility can kept intact.The following table shows the sugarcane production of Maharashtra. Sugarcane is also identified as one of the major cash crop.

Table 1: Sugar production of Maharashtra (Courtesy: indianexpress.com/article/india)

Production Year	*Production of Sugar (Lakhs Ton)*		*Recovery of sugar(percentage of sugarcane crushed)*	
	MH	*UP*	*MH*	*UP*
2011-12	89.96	69.74	11.67	9.07
2012-13	79.87	74.85	11.41	9.18
2013-14	77.12	64.95	11.41	9.26
2014-15	105.14	71.01	11.30	9.54
2015-16	84.15	68.55	11. 33	10.62
2016-17	42.00	87.73	11.26	10.61
2017-18	107.21	120.50	11.24	10.84
2018-19	98.45	81.77	11.14	11.27

From the above table we get to know that India is producing Sugarcane as the highest contribution to national income. Also the recovery of sugar is also high for Maharashtra.

II. LITERATURE SURVEY [2, 3]

Table 2: Literature survey

Author and publication	*Techniques applied*	*Variables considered*	*Area o grow*
Saeed Khaki,2019	deep neural network	genotype, environment,	black box property
Mrs.K.R.Sri Preethaa,2018	Bayesian algorithm	Parameters of soil, fertilizer used, duration of crop and humidity	Suggestion of pesticide for different type of disease
Dr.A.Senthil Kumar, P.Arun,2017	K nearest neighbor, Artificial neural network and data mining techniques	Comparative study	-

E. Manjula,2017	data mining technique based on association rules	Year, District, Crop, Area, Tanks, Bore Wells, Open Wells,	Need to check with other algorithm
Sujatha , 2016	SVM,ANN, random forest, J48, Naïve Bayes Algorithms are used	Static weather Parameter and crop attributes are used for crop prediction	Real time data and number of variables need to consider

III. METHOD

6 completed crop cycles records from 80 fields are taken into consideration. These records are extensively edited to provide figures appropriate for statistical analysis and electronic manipulation. Observations that are not representative were disqualified from the exploration using 3 prior criterion:

a. elimination of all observations with no fertilizer (N,P,K) records;

b. rejection of records which differ by more than 10 % duplication.

c. Excluded observation made on mixed varieties of sugarcane which contribute less than 15 records.

The experimental data from 80 farmers were collected. The data cleaning and analysis is performed on data. Missing values are filled using mean method. Some observations are ignored as its giving very high error due to scattered of data.

The method of least squares is used for linear regression analysis. It is used to found a functional relation between the variables. The model which is generated using the formula

Regression Line = Minimize $(\sum (Y_i) - (\beta_0\text{^} + \beta_1\text{^} x_i))^2$

Where we need to get function which will generate low bias.

The dependent parameter i.e. Y which is yield in our multiple regression model. The independent variables are values of N, P, and K fertilizer. The model generated on three season dataset.

We found in most of the model N and P parameters are affecting yield parameter in the model as compared to K parameter. As we tried to generate the model by considering more features or parameters but due to very few records are available and which is scattered also it gives very high error. Due to this the error generated is very high and prediction is not up to the mark. So we have not considered maximum parameters but limit the dataset.

IV. EXPERIMENT AND RESULTS

A. Multiple Regression Model for Adasali Sugarcane

Adasali sugarcane is 18 months crop. Its sowing time is till June to August. The Multiple regression method will consider one dependant variable that is crop yield and remaining variables such as N, P, K, water variables, planting densities, energy usage.

After applying the multiple regressions on train data set we will get the following Model. We will only consider coefficients

Table 3: Multiple Regression Model for Adasali

Parameters	*Coefficients*	*P-value*
Intercept	85.50959	0.001555
N	0.149396	0.000226
P	0.176014	0.043106
K	0.178363	0.010977

And intercept for building a model. The P value of the variable is greater than 0.15 is not considered for prediction. So according to result at Table 1 our Model is as shown below

Equation 1:

Predicted Y = 85.50959232+0.149396062* value of X1 0.176013688*value of X2+ 0.178363252*value of X3

For the problem in linear regression we try to find the predicted value of Y.

$$Y = \beta_0 + \beta_1 X$$

To find the values for the coefficients i.e. β0^ and β1^ that minimize the objective function we take the partial derivates of the objective function (SSE) with respect to the coefficients. Set these to 0, and solve.

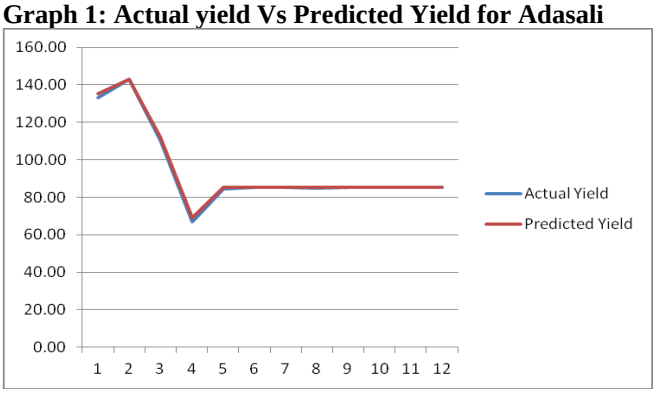
$$\beta_1 = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - \left(\sum x\right)^2} \qquad \beta_0 = \frac{\sum y - \beta_1 \sum x}{n}$$

For multiple regression for n multiple variables we have formula,

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n$$

Where β_0 is constant which is intercept in our formula, X1,X2,…Xn are the values of N,P, K from the dataset and $\beta_{1,\ldots}$ B_n are the slopes used for getting predicted value Y.

Graph 1: Actual yield Vs Predicted Yield for Adasali



So for our model the error of prediction is 12.52.

The square of error is 1.043333333.

B. Multiple Regression Model for Pre-Seasonal Sugarcane

Preseason Sugarcane is of 16 months crop. Its sowing time is from October to November. After performing multiple regressions analysis on the data gathered from farmer. We got the following calculation

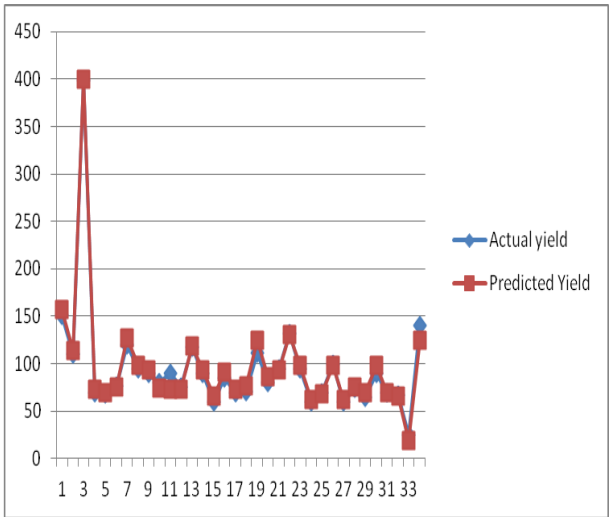
From the table the Model we get as follows, Equation 3

Predicted Y = -

31.2418932+0.663990431*value of Variable N - 0.77497584*value of Variable P+ 1.144635649*value of Variable K

Table 4: Multiple Regression Model for Preseasonal

Parameters	*Coefficients*	*P-Value*
Intercept	-31.2418932	0.68513154
N	0.663990431	0.023323571
P	-0.77497584	0.036496118
K	1.144635649	0.047619227



Graph 2: Actual yield Vs Predicted Yield for Preseasonal sugarcane

So the error of prediction for our model is -47 and the standard error of estimate is -2.043478261.

C. Multiple regression Model for Suru/Seasonal Sugarcane Suru or Seasonal sugarcane

Crop duration is of 12 months. Sowing time for crop is January to February. The following is the result obtained for Multiple Regression on the data.

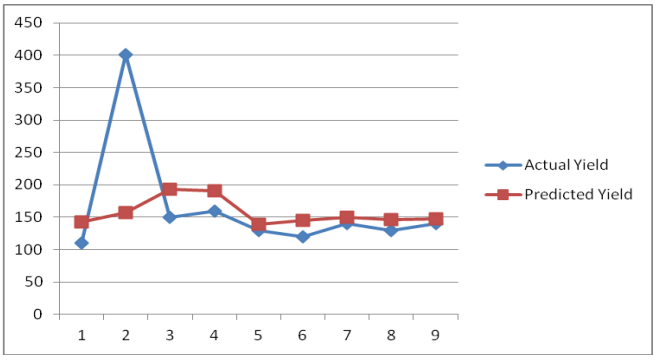
Table 5: Multiple Regression Model for Suru

Parameters	*Coefficients*	*P-value*
Intercept	190.5184576	0.000317
N	-0.121734273	0.564733
P	-0.193451256	0.568065
K	0.013255739	0.973154

So our model of regession is Equation 4

Y Pred =190.5184576+N*0.121734273+P*0.193451256 + K*0.013255739

So the error of prediction for our model is 76.71594 and the standard error of estimate is 7.671594.

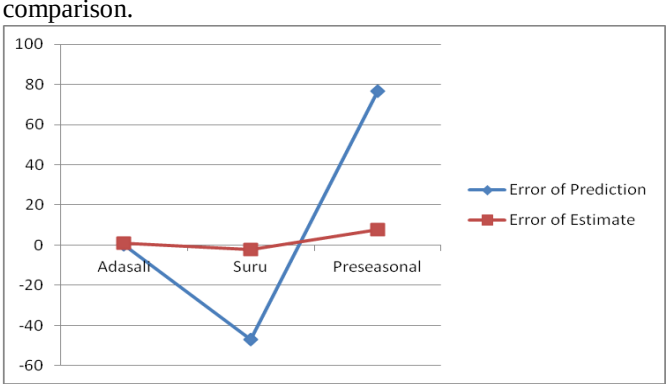


Graph 3: Actual yield Vs Predicted Yield for suru sugarcane

From the models that we obtained states that the crop yield is mostly depend on the parameters which are N, P, K and water resource. The quantity of the fertilizer mainly N, P, K can be applied to the soil can also suggested to the famer depending on the available soil N, P, K parameters.

The soil health card issued by government can be used to get the N, P, K values.[4]

From the Models, we got the error of estimate and error of prediction for each data set. The graph below shows the comparison.



Graph 4: Error of prediction and estimate for three season

As from the graph it shows that the regression model of season Suru gives low error of prediction and estimate. This shows that the Crop prediction given by this model is more accurate compared to other models.

V. FERTILIZER RECOMMENDATION DEPENDING ON SOIL PARAMETER

The equations which can be used for determination for fertilizer recommendation are as follows:

Table 6 :Suru Hungama(WITHOUT COWDUNG FERTILIZER)	
N	4.76*Production expectation-1.34*AVILABLE N OF SOIL
P	1.24*Production expectation-1.55*AVILABLE P OF SOIL
K	2.73*Production expectation-0.21*AVILABLE K OF SOIL

Table 7 :Purva Hungama (WITHOUT COWDUNG FERTILIZER)	
N	3.79*Production expectation-1.16*AVILABLE N OF SOIL
P	1.53*Production expectation-2.61*AVILABLE P OF SOIL
K	3.67*Production expectation-0.73*AVILABLE K OF SOIL

Table 8: Khodava (WITH COWDUNG FERTILIZER)	
N	3.89*Production expectation-0.94*AVILABLE N OF SOIL-0.94*cowdung Fertilizer
P	1.12*Production expectation-1.66*AVILABLE P OF SOIL-40* cowdung Fertilizer
K	3.06*Production expectation-0.58*AVILABLE K OF SOIL-1.04 *cowdung Fertilizer

Table 9 :Khodava (WITHOUT COWDUNG FERTILIZER)	
N	4.47*Production expectation-1.08*AVILABLE N OF SOIL
P	1.56*Production expectation-2.32*AVILABLE P OF SOIL
K	3.37*Production expectation-0.64*AVILABLE K OF SOIL

Table 10 : Purva Hungama (COWDUNG FERTILIZER)	
N	3.73*Production expectation-1.14*AVILABLE N OF SOIL-3.09*cowdung Fertilizer
P	1.49*Production expectation-2.54*AVILABLE P OF SOIL-2.24* cowdung Fertilizer
K	3.15*Production expectation-0.63*AVILABLE K OF SOIL-1.16 *cowdung Fertilizer

Table 11 : Adasali Hungama (WITHOUT COWDUNG FERTILIZER)	
N	4.39*Production expectation-1.56*AVILABLE N OF SOIL
P	1.62*Production expectation-4.56*AVILABLE P OF SOIL
K	1.86*Production expectation-0.37*AVILABLE K OF SOIL

VI. OBSERVATION

Following are the observations of models generated for Adasali, Suru and preseasonal.

Firstly the dependent variable in the model is sugarcane yield which basically depends on N, P, K and other variable whereas water resource need to be treated important, but sometime it will behave unexpectedly as we know that agriculture is very uncertain branch. So when to build model on area like agriculture need to consider each and every minute parameter to reach to the accuracy expected , but it again depend on the real parameters and accurate static data. Data from Indian government on agriculture will really help the researcher to contribute for farming through technologies. Secondly, sometime many of the parameters considered are really independent on each other and which surely affects the performance of the model. So simple multiple regression is not going to give the desired result as well as accuracy for the output variable. The output of our model is generally helpful to managers for planning and resource allotment required for achieving certain desired yield. The appropriate and ample amount of data and technologies like neural network along with function activation and weight adjustment may solve the purpose.

This may give the desired result.

VII. VALIDATION

For predicting future sugarcane production there is need of the sugarcane crop dataset at least 6 to 8 months in advance to take decision on parameters. This is required to analyze and visualize the relationship with each variable. From the models generated the results for yield prediction does not fit appropriate for all the future data.

The accuracy of the model is very low while error generated is very high. So model generated with multiple regressions using available data is very poor.

The cross validation method might help to improve model accuracy and may give better result for unseen data.

VIII. CONCLUSION

The area of study on which the model is applied is restricted to arid region of Maharashtra; Pandharpur.The model is trained and tested with the input values given by 80 farmers. The model can be undertaken for more number of records and need to train accordingly.

The model recognized and executed focuses on parameters which were treated important initially. The supplementary research is required to find out the association and patterns between dependent and independent variable.

ACKNOWLEDGEMENT

The Annual diary “Oos Sheti Dnyanayag” published by Vasant Dada Sugar Institute (VSI) Research center .VSI is recognized as a Centre for Research by the Department of Science and Technology of the Central Government. Research and publication by VSI really helped to reach the conclusion so the research is gratefully acknowledged.

REFERENCES

1. EA brüggemannac , JR kluga, PL greenfielda AND HM dicksb ,“Emperical Modelling and Prediction of Sugarcane yields from field records”
2. Mr. Saeed Khaki 1; lizhi Wang 2., “Crop Yield Prediction Using Deep Neural Networks”. By Frontiers 2019.
3. Mrs. Shivani S. Kale and Dr. Preeti Patil , “Data mining technology with fuzzy logic ,neural networks and machine learning for agriculture”, at ASIC book series “Data management ,analytics and innovation by springer 2018.
4. Mr. V. Sellam and E. Poovammal ,”Prediction of crop yield using regression analysis”,Indian Journal of Science and technology,Vol 9 (38),October 2016.
5. Mrs. Shivani S. Kale and Dr. Preeti S. Patil, ”Use of data mining technology in agriculture sustainable development”, IJCRT, Volume 6, Issue 2. April 2018.

AUTHORS PROFIL

Shivani S. Kale, Assistant professor KIT, Kolhapur, Research scholar VTU Belgaum Karnataka, CSI Member,01-Indian Patent, Paper Published at Springer and other Journals.

Dr. Preeti S. Patil, Designation HOD and Associate Professor. International Journals- 15 • International Conference- 18 • National conference- 43. ISTE (LM), CSI (LM). Book Published ISBN: 978-93-5164-482-8, Patents Published.