International Journal of Advance Engineering and Research Development
Scientific Journal of Impact Factor (SJIF): 4.72
Special Issue SIEICON-2017,April -2017
e-ISSN : 2348-4470 p-ISSN : 2348-6406

CROP YIELDING PREDICTION APPLYING REGRESSION: FORECASTING WHEAT/RICE YIELD FOR ANAND DISTRICT

Saniya Vhora[1], Ankita Gohil[2], Priya Shah[3], Bhagirath Prajapati[4], Priyanka Puvar[5]

[1.] **Department of Computer Engineering, A.D. Patel Institute of Technology, Gujarat, India**
[2.] **Department of Computer Engineering, A.D. Patel Institute of Technology, Gujarat, India**
[3.] **Department of Computer Engineering, A.D. Patel Institute of Technology, Gujarat, India**
[4.] **Department of Computer Engineering, A.D. Patel Institute of Technology, Gujarat, India**
[5.] **Department of Computer Engineering, A.D. Patel Institute of Technology, Gujarat, India**

***Abstract*—*Agrarian territory in India is facing drastic problem to maximize the crop productivity. The problem of yield prediction is a primary issue that remains to be solved based on receivable statistics. A new later evolution in Information Technology for agriculture zone has become a charming research area to estimate the crop turn-out. Varied data Mining techniques are utilized and appraised in agriculture for predicting the forthcoming year's crop manufacture. Data mining techniques are the better preference for this objective. The environmental parameters like rainfall, sunlight, evaporation, Humidity, Temperature etc. that impacts the yield of crop and to implant relationship among these parameters. Yield prediction boots the peasants in impairing the losses and to get best prices for the crops. Estimation of food grain production credible omnibus and timely info on the food situation may patronize to the government policies. This paper represents a concise analysis of crop yield prediction using Multiple Linear Regression (MLR) technique for the selected domain. By applying Regression and finding the correlation of each parameter with Yield we are getting our Regression Model that gives 0.95(95%) value of R Square (Coefficient of Determination).*

Keywords- *Multiple Linear Regression, Statistics Approaches, Data Mining,Rabi wheat , weather indices,Coefficient of Determination.*

I. INTRODUCTION

Prediction of crop yield is a frenzied challenge. The economy of India is largely agricultural. Forecasting crop yield well before harvest is crucial especially in regions characterized by climatic uncertainties. This enables planners and decision makers to predict how much to import in case of shortfall or optionally, to export in case of surplus. It also enables governments to put in place strategic contingency plans for redistribution of food during times of famine. Therefore, monitoring of crop development and of crop growth, and early yield prediction are generally important. Improved understanding of the effects of climate change on crop yield is central to planning appropriate responses. Analysts wishing to anticipate these effects must inevitably rely on some numerical model of how crop yields respond to climate **[5].** A common approach is to use statistical model trained on historical yields and some simplified measurements of weather, such as Annual Rainfall, sunlight hours, precipitation, Evapotranspiration, seasonal temperature, wind speed, Humidity etc. The performance of statistical models differed by ambiance variable and spatial scale, therefore statistical models are likely to continue to play an important role in anticipating future impacts of climate change. Statistical models, in which historical data on crop yields and weather are used to calibrate relatively simple regression equations, provide a common alternative to process based model. Statistical models are not without serious shortcomings, however, and in particular, they are subject to problems of co-linearity between predictor variables (e.g., temperature and precipitation, sunlight hours, Rainfall), assumptions of stationary, those past relationships will hold in the future.

Regression Analysis can be defined as a structured approach which stresses on the analysis of data for the research purpose on decision making and problem solving **[2].** Regression Analysis is a commonly used technique in the research where relationship among the considered variables has to be established and to identify their effects on crop yield. Crop yield is considered as a dependent variable and weathers as independent variables **[1].**

The crops can be considered for analysis are Rice and Wheat because they are the most common crop cultivated in many areas of India. For this project, we are going to analyze and predict the crop yield for the Anand District in Gujarat.

II. PREVIOUS WORK

The experiment of agricultural sustainability has grown into more extreme in recent years with the sharp rise in the cost of food and energy, climate change, water scarcity, degradation of natural ecosystems and biodiversity, the financial crisis and expected increase in population. Researchers who are doing some enhancements in the Agricultural meteorology, predicting the crop yield on the basis of the region's climatic data. And then finding how much the climatic parameter will respond. The research on this prediction modeling topic is done by statistical approaches, Artificial Networks, cluster

analysis and classification techniques. For this project of Estimating the yield of the crop will be done by applying the statistical method, Multiple Linear Regression. Researchers have forecast or predicted the crop Yield of many crops like Wheat, Rice, Mustard, and Tobacco and for so many crops.The Regression Model then generated should have less than 0.05 significant value and should have at least 0.85(85%) value of R-Square (the Coefficient of Determination.) and Multiple R (The correlation coefficient). The results show that climatic parameters influenced the yield of crops differently in different districts.

III. METHODS AND MATERIALS

A. Numerical Resources (Database)*:*

The numerical resources or the historical data for the yield prediction or yield estimation is of Anand District of Gujarat state. This database is meteorological conditions or climatic data of Anand. Improved understanding of the effects of climate change on crop yield is central to planning appropriate responses. The weather data of Anand District contains weather parameters like Rainfall, sunlight hours, precipitation, evapotranspiration, seasonal average temperature. First of all by applying Multiple Linear Regression on these weather parameters as Independent variables and Yield as the Dependent variable we are supposed to find significant parameters in which parameters are mainly contributes in Yield estimation. And hence after applying Multiple Linear Regression on these historical data observe the Regression Analysis output sheet and its different output terms and Estimating the yield **[6].**

B. Statistical Methods:

There are numerous methods of inward at forecasts of production of food crops. These are: (a) Observing crop circumstances on the basis of agro-climatic data, (b) building regular survey to assess the area, yield and production of crops; and (c) estimating regression models describing the quantifiable relationship between selected climate/input variables and the final yield of the crop. And how crop yield will respond to weather? That we can infer.

A better approach is to use statistical model trained on historical yields and some simplified measurements of climatic parameters Statistics are the discipline of gathering, describing and scrutinizing data to quantify variation and uncover useful relationships **[5].** Whether your goal is description, prediction or explanation, you will escalate the statistical discovery paradigm. In data mining, statistical approaches are used in prediction modeling.

C. Data Mining in Prediction Modeling:

Data Mining is the analysis step of the "knowledge discovery in databases". The term is a mislabeling or misnomer because the aim is the extraction of patterns and acquaintance from large amounts of *data***,** not the extraction **(***mining***)** of *data* itself.Predictive modeling is a process that uses Data-Mining and probability to estimate outcomes. Each model is made up of a number of predictors, which are variables that are likely to manipulate future results. Once data has been collected for relevant predictors, a statistical model is formulated **[2].** The model may employ a simple linear equation or it may be a complex neural network, mapped out by sophisticated software. As additional data becomes available, the statistical analysis model is validated or revised. Predictive modeling is frequently allied with Meteorology and Weather forecasting.Analyzing representative portions of the available information (sampling) can help pace development time on models and allow them to be deployed more hastily. And hence Prediction or Estimation is having main importance in modeling and Analyzing in Agricultural Field **[1].**

D. Multiple Linear Regression:

Regression analysis is used to analyze and determine the relationship between the response variable and explanatory variable. Multiple linear regression attempts to model the relationship between two or more explanatory variables and a response variable by fitting a linear equation to observed data. The variables considered for analysis in this research work are Rainfall, Area under Cultivation, Evapotranspiration, Sunlight hours, seasonal Temperature, soil type. Crop yield is a dependent variable which depends on all these organic factors. We need to find significant parameters or we can say that checking for significances for these climatic parameters and then we can accomplish from that which parameters are mainly contributing in Yield prediction.For this project, we are supposed to apply MLR on this weather data in Microsoft Excel using Analysis Toolkit **[4].**And from these Historical data of Anand districts climatic datawe need to predict how much yield the farmer will about to expect and from best prediction analysis he can change his decision or can add any factor that yield will need. Formally, the model for Multiple Linear Regression, given *n* observations:

$$\mathbf{Y_i = \beta 0 + \beta_1 x_1 + \beta 2 x 2 + \ldots \ \beta_n x_n.}$$

For *i* =1, 2,…n.

Here, Y is Dependent Variable

β0, β1, β2…. βn are the Coefficients

X1, X2,…Xn are the Independent Variables

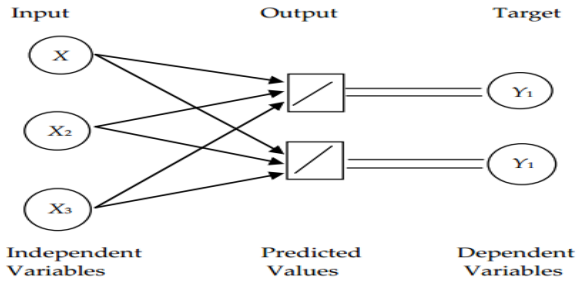

Figure.1Multiple Linear Regression

IV. PROPOSED WORK

Estimating the Crop Yield benefits the farmers in reducing their losses and to get best prices for their crops. Sector of Agriculture in India is facing meticulous problem to exploit the crop productivity.The problem of yield prediction is a main or key problem that remains to be solved based on available historical data. As mentioned earlier the climate data in use for the research is of Anand District. First of all, we need to find significant parameter, which will be used or contributes to achieving more Yield production. After finding the significance of each parameter, apply MLR between the percentage of maximum error parameters (Independent variables) and the Crop Yield (Dependent variable)[7].Then by observing the regression output sheet we can conclude to our work and find the best possible way to estimate or predicting the more production of crops yield. The explicit formulation of the multiple regression models is:

$$Y= \beta_o +\beta_1(RF) +\beta_2(BSS)+\beta_3(MAXT)+\beta_4(MINT) + \beta_5(RH) + \beta_6(WS) \ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

Where,

Y : Yield of Crop (kg/ha).
BSS : Bright Sunshine hours (hr)
ET : Evapotranspirasion (mm/ha)
RH : Relative Humidity
WD : Wind Direction
WS : Wind Speed
RF : Rain Fall (mm)
MAXT: Maximum Temperature (Degree C)
MINT : Minimum Temperature (Degree C)
VP : Vapor Pressure (Bar)

V. DATASET

The dataset taken under consideration for the Analysis And for the Estimation of the crop's Yield is of Anand District. Anand district in Gujarat is most cultivated district and the main crops harvested here are mainly Wheat and Rice. The Weather data we have is from 1958 but due to some missing values and attributes, we have considered the data from 1985 to 2016.From this dataset, we have found the significant variables or we can say that attribute which has more contributed to getting more and accurate predicted Yield. This dataset will always be there in the background for user's understanding and for further considerations. Data for the prediction is from 1985 to 2012 and Data from 2012 to 2015 are chosen for the validation.

A. Weather variables:

The tactic suggested by Indian Agricultural Statistical Research Institute (IASRI), New Delhi the crop yield forecasting models for major growing districts were developed using stepwise regression analysis.Climate variables used for the model are bright sunshine hours (BSS), maximum temperature (Tmax), minimum temperature (Tmin), morning relative humidity (RH1), afternoon relative humidity (RH2), morning vapor pressure (VP1) and afternoon vapor pressure (VP2) **[3].** Table.1 shows, datasets from 1985 to 2016 having different climatic parameters as Independent variables week wise.They use the best available science, as well as three key variables that are vital to perceptive weather: air pressure, temperature and air density. These variables are necessary because, like a well-organized set of drill sergeants, they control how air behaves, and thus, they control the weather, thus, Weather parameters are must need for the prediction process of Yield.

week	EP	BSS	RF	WD1	WD2	WS	MAXT	MINT	MeanT	RH1	RH2
1	2.6	7.5	0.8	3.9	8.9	3.6	26.3	12.1	19.2	96	50
2	2.9	9.2	0.0	1.4	12.7	1.4	29.8	9.2	19.5	99	36
3	2.9	8.8	0.0	2.9	14.0	3.0	26.9	11.9	19.4	88	58
4	2.8	6.1	1.0	2.9	8.9	3.3	25.7	12.6	19.1	93	53
5	3.2	9.8	0.0	3.6	12.0	2.6	28.2	10.5	19.4	92	39
6	3.8	9.3	0.0	2.9	11.4	3.4	29.4	13.4	21.4	85	44
7	4.3	9.9	0.0	0.7	16.6	2.3	31.8	12.7	22.2	90	36
8	4.5	9.7	0.0	15.1	24.3	2.3	35.3	16.3	25.8	87	36
9	4.2	8.1	25.4	10.4	19.1	3.7	27.8	14.0	20.9	89	50
10	5.2	9.7	0.0	5.3	12.7	3.1	32.2	16.0	24.1	82	34
11	5.6	9.0	3.6	14.0	16.6	3.7	32.8	16.8	24.8	82	42
12	7.0	10.3	0.0	14.6	23.0	2.4	38.3	19.0	28.6	72	23
13	6.7	8.3	0.0	18.4	23.0	3.4	38.7	21.6	30.2	77	34
14	7.4	8.7	0.0	20.4	23.0	4.4	36.2	21.5	28.9	82	43
15	5.9	9.2	18.8	15.3	17.9	4.1	36.0	21.8	28.9	90	45
16	7.4	10.2	0.0	14.0	28.1	3.1	40.1	23.4	31.8	74	24

Table: 1 Dataset (Weekly)format for the Analysis

VI. RESULTS AND DISCUSSION

Agriculture is extremely dependent on ambiance and, as such, crop yield changeability is affected by year-to-year climatic changeability, with regards to both tremendous events and changes in historical patterns of regional climate. This vulnerability could have noteworthy effects on crop production because of the many qualms for the growing season and may result in financially viable and food security risks in some parts of the world.The final models were selected on the basis of highest R2 and the value of significance of F test.For validation of the model Three years actual yield data viz. 2012-13, 2013-2014 and 2015-16 were used **[3].**

A. Preparing Weather Indices:

- Arrange weekly weather data of Tmax, Tmin, Rain, RHI and RHII for at least 25 years.
- Indices are nothing but all the possible combinations of all the weather parameters by doing addition and then doing sum-product of the particular combination.
- Our first combination is BSS-Tmax, then for finding first Indices, do the sum of all weekly BSS and Tmax and then for Second Indices, Do sum-product of all weekly BSS and Tmax values.
- Likewise, find all the possible Weather Indices which are, total 56 Indices.
- Then, perform Multiple Regression on all the 56 Indices as Independent variables and Yield as the Dependent variable.
- Hence, As a result, we are getting R-square value (coefficient of determination) as 0.95(95%) for Wheat and 0.89(89%) for Rice.
- Refer, Table: 2, The Results, Predicted Yield for 2016-2017, we are getting from the developed Regression Equation.

Yield forecast models of Wheat and Rice crop for Anand district of Gujarat developed using composite weather variables

Sr.No.	Crop	Regression Equation	R-Square	Forecast Yield(Kg/ha
1.	Wheat	Y=2018.323+(53.509*Time)+(1.065* MINT*RH2)+(-0.661*BSS*VP1)	0.95	3194.217394
2.	Rice	Y=12260.253+(60.331*Time)+(0.060* RF*RH1)+(-21.895*MINT)	0.89	2847.572

Table: 2 From Regression output the Regression Equations we are getting.

B. Results:

The crop yield forecasting models were developed using the weather data up to 2016 at pre- harvest season stage for one of major growing district Anand in Gujarat state.
The coefficient of determination (R-Square) was ranged 0.95 for Wheat and 0.88 for Rice. The maximum wheat yield was predicted for Anand district (3194.217394 kg/ha) and the maximum Rice yield was predicted for Anand district

(2847.57 kg/ha).The maximum errorpercentage of wheat crop during 2015-16 was observed in Anand district with -8.01741 and for Rice with -1.2 percent of observed.

We have validated that the predicted Yield for the year 2015-2016 is 2721.61 kg/ha and the actual Yield is 2789.67 kg/ha and due to the minor error this predicted result is accurate and reasonable. The main Aim for this project research is to generate appropriate Regression model and predicting the accurate Yield for the major crops that are Wheat, Rice, and Mustard and so on. And hence the Estimated Yield for Wheat in 2016-2017 is 3194.217394 kg/ha from the historical pattern and Regression Analysis. Refer Table: 3, that shows all the values of actual Yield and year wise appropriate Predicted Yield.

For the 2016-2017 Year, the predicted Yield from the generated Regression Equation is 3194.217394 kg/ha.

YEAR	ACTUALL YIELD	PREDICTED YIELD	YEAR	ACTUALL YIELD	PREDICTED YIELD
1985-86	1688	1517.163563	2001-02	2372	2358.637316
1986-87	1365	1483.914012	2002-03	2413	2516.949805
1987-88	1894	1917.015027	2003-04	1937	1894.816888
1988-89	1611	1685.470321	2004-05	2674	2675.267164
1989-90	1715	1642.787286	2004-06	2610	2705.976449
1990-91	1526	1489.165696	2004-07	3079	2639.822459
1991-92	1814	1686.571139	2004-08	2942	2778.117885
1992-93	2084	1880.576007	2004-09	3239	3175.434493
1993-94	1656	1676.688494	2004-10	2535	2881.464069
1994-95	1794	1867.372897	2004-11	2655	2573.984838
1995-96	1981	2070.982951	2004-12	2558	2679.279449
1996-97	1798	1983.244154	2004-13	2822	2704.706176
1997-98	1959	1893.899398	2013-14	2585	2792.250005
1998-99	1694	1914.431926	2014-15	2932	3047.684636
99-2000	2079	2195.165203	2015-16	2852	2721.618676
2000-01	2090	2095.278197	2016-17		3194.217394

Table: 3 Yearly Predicted Yield

A. How the association of predicted yield is done with old data presented?

From the Table 3, we have predicted the Yield of Wheat Crop by using the generated Regression Equation (refer Table 2). The Actual Yield shows the real Yield production of each year. We have collected the climatic data and year wise produced Yield of wheat crop from 1985 to 2016 from Anand Agricultural University, Department of Meteorology. Then By developing Regression Equation we have found the predicted Yield for each year from 1985 to 2016.And hence we can observe and then compare the values of Actual yield and predicted yield having minor error and minor difference.

VII. ANALYSIS

The study was carried out to forecast the yield of wheat, and Rice of Anand district in middle Gujarat agro-climatic zone of Gujarat. The daily weather data viz. maximum temperature, minimum temperature, morning and afternoon relative humidity, morning and afternoon vapor pressure were arranged week-wise from showing to the physiological development and the relation between weather parameters and yield was resolute using statistical tools like correlation and Regression Analysis. For Wheat, data is arranged weekly from 44th week to 52nd and 1st to 7th week of every year (1985-2016) and for Rice data is arranged weekly from 22nd week to 41st week of the year.

A. How the past yield data affects the prediction?

As we know that, Regression Analysis is the most widely used statistical approach for prediction modeling. Here, for our project, predicting the crop yield from the past Historical weather data and yield data and then by developing Regression Equation we are getting the predicted Yield. Hence, that is how the past yield data and will used to predict future yield and it will used for validation too.

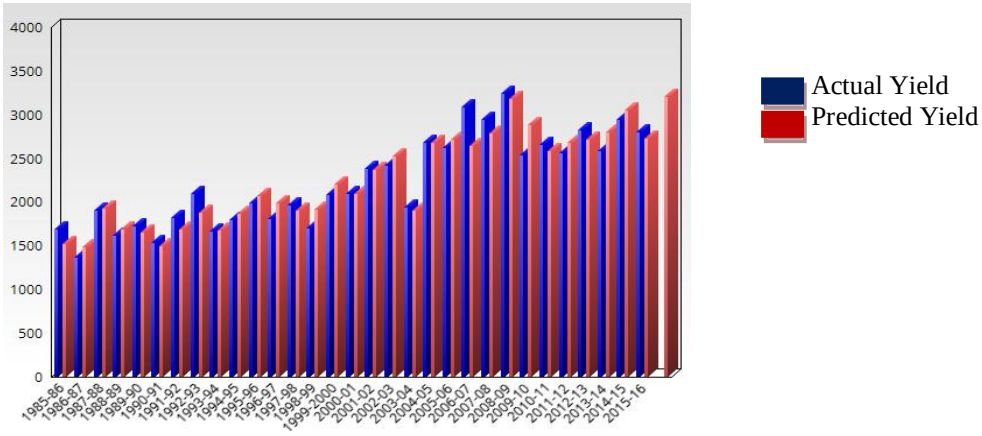

Figure.1 Yield Vs. Year Graph

VIII. CONCLUSION

Results from the studies show that the statistical models based upon the weather indices are successfully simulating pre harvest yield forecast of wheat and Rice crop.This model is simple, does not require any sophisticated statistical tools,required only climate data for crop rising periods, past yield data and provides the good pre-harvest forecast. Therefore it can be used for district, agro-climatic zone and state level forecast. By performing Multiple Linear Regression variables used the Regression Equations are developed from the values of the coefficients and Hence, Yield for the Wheat and Rice crops will be predicted or Estimated for the year 2016-2017 for Anand District in Gujarat.

IX. REFERENCES

[1].V. Sellam and E. Poovammal. "Prediction of crop Yield using Regression Analysis." Indian Journal of Science and Technology, October 2016.

[2].M. Sayedur Rahman,Md. Monimul Huq,M.and M. Raque Azad." Statistical Analysis of Crop-Weather Regression Model for Forecasting Production Impact of Aus Rice in Bangladesh". International Journal of Statistical Sciences,2005.

[3]. Bruno Basso, Davide Cammarano, Elisabetta Carfagna."Review of Crop Yield Forecasting Methods and Early Warning Systems".Department of Geological Sciences Michigan State University, USA.

[4].D. D. Patil, Rucha Dave, S. B. Yadav and Vyas Pandey."Yield forecasting of wheat, mustard and potato crops by the statistical model of different Districts in Gujarat".Department of Agricultural Meteorology, B A College of Agriculture, Anand.

[5].Young Mature Oil Palm ,Yong Kian Keong." Statistical Modeling of Weather-based Yield Forecasting for Young Mature Oil Palm".ICAAA, July 23-24, 2012, Singapore.

[6]. Louis M. Thompson. "Multiple Regression Techniques in The Evaluation Of Weather And Technology In Crop Production". Associate Dean of Agriculture I Iowa State University.

[7]. Ananta Vashisth, R. Singh, Manu Choudary." Crop Yield Forecast at Different Growth Stage of Wheat Crop using Statistical Model under Semi Arid Region" Division of Agricultural Physics Indian Agricultural Research Institute, New Delhi. Journal of Agroecology and Natural Resource Management.

