Journal of Green Engineering (JGE)

Volume-10, Issue-3, March 2020

Analysis of Ecological Parameter for Predicting the Crop Yield Using Multivariate Regression

[1]V.Sellam, [2]N.Kannan, [3]E.Poovammal

[1]Assistant professor,Department of Computer Science And Engineering,SRMIST-Ramapuram,Tamil Nadu,India. E-mail: sellamv@srmist.edu.in
[2]Professor and Head, Department of Computer Science And Engineering, SRMIST-Ramapuram,Tamil Nadu,India. E-mail:n.kannan.k@gmail.com
[3]Professor, Department of Computer Science And Engineering, SRMIST-Kattankulathur,Tamil Nadu,India. E-mail: poovamme@srmist.edu.in

Abstract

Forecasting in agriculture is otherwise known as foreseeing the production of a chosen crop. Foreseeing the production is not easy because the constraints and parameters involved are uncontrollable beyond certain point. The agricultural output is highly unpredictable due to various reasons in recent times. Crop yield forecasts are important for advance planning, formulation and implementation of policies related to the crop procurement, distribution, price structure and import export decisions etc. The parameters considered are Minimum Support Price (MSP), Consumer Price Index (CPI), Food Price Index (FPI), Annual Rainfall (AR) and Area under Cultivation (AUC).The Crops considered for analysis is Rice, Maize and Gram. This foreseeing not only help the farmers but also helps the policy makers and agencies to plan in regard with crop procurement, import and export and decisions on agriculture based industries. The influence of economic factors (MSP, CPI and FPI) and ecological factors (AR, AUC) on CY is analyzed using Multivariate Analytical techniques (MVA).In this work an attempt is made to predict the Crop Yield using Multivariate Regression.

Keywords: Crop Yield Analysis, Ecological Parameter, Socio-Economic impacts, Price Indices, Prediction

Journal of Green Engineering, Vol. 10_3, 593-614.
© 2020 *Alpha Publishers. All rights reserved*

1 Introduction

India is an agricultural country where majority of the population is mainly involved in agriculture. India is proud to have plenty of natural wealth which includes fertile lands, forests, mountains and rivers. Annual Rainfall due to torrential rains also plays an important role in Indian agriculture. Climate changes due to global warming and other factors radically changes the seasonal rainfall and in-turn affects the Crop Yield.

Advanced water conservation and water management techniques including advanced irrigation techniques and recycling of water should be developed to increase the Crop Yield. India has very good water resources for irrigation and underground water level is also far better than other countries but most of the river water is not properly utilized and goes to the sea. Developed countries like UK and USA have made a total green revolution adopting the best water conservation techniques; water recycling techniques and advanced machineries for its agriculture.8% of global agricultural gross domestic product is contributed by Indian agriculture to support 18% of the world population. Hence agricultural research using data mining techniques helps the farmers in a big way.

The overall yield of a region depends more on the factors like extreme weather conditions, unpredictable rainfalls during planting or harvesting, soil type and changes in product prices. The yield of a crop is affected by many financial indicators that reflect the economic value of agricultural outputs such as Minimum Support Price (MSP), Consumer Price Index (CPI) and Food Price Index (FPI). MSP is the minimum price fixed by the Government for procuring products from farmers by the agencies or government, thereby preventing the distress sale leading to heavy losses for farmers. Consumer Price Index expresses the difference in the cost of goods and services between the current year and the previous year to show the effect of inflation on purchasing power. The measure of the monthly changes in international prices of a set of food commodities is represented by FPI.

Numerous agricultural applications and total workforce involvement are analyzed using various pattern classification techniques like SVM, K-means [1]. It also deals about the need of analyzing data in the field of agriculture using different models. To increase the yield of crops, potential and prospective cropping areas are identified using Relative Spread Index (RSI) and Relative Yield Index (RYI) [2].

The farmers are required to take proper decision in choosing the crop from time to time depending on the water availability and market demand. The yield of crops generally depends on rainfall distribution, irrigation pattern,soil nutrition, physical, ecological and social characteristics.

The Fig.1 illustrates the factors affecting the maximum crop yield and selection of crops seasonal based. The uncertainties of the factors are taken into account to optimize the decision parameters for analysing the impact on farmers' returns in terms of yield and price. Cultivation of crops involve several decisions which includes selection of crop variety, allocation of land, planting, harvesting, storing, and selling. Before the sowing season, farmers will not have exact information on weather, yield, price, supply and demand.

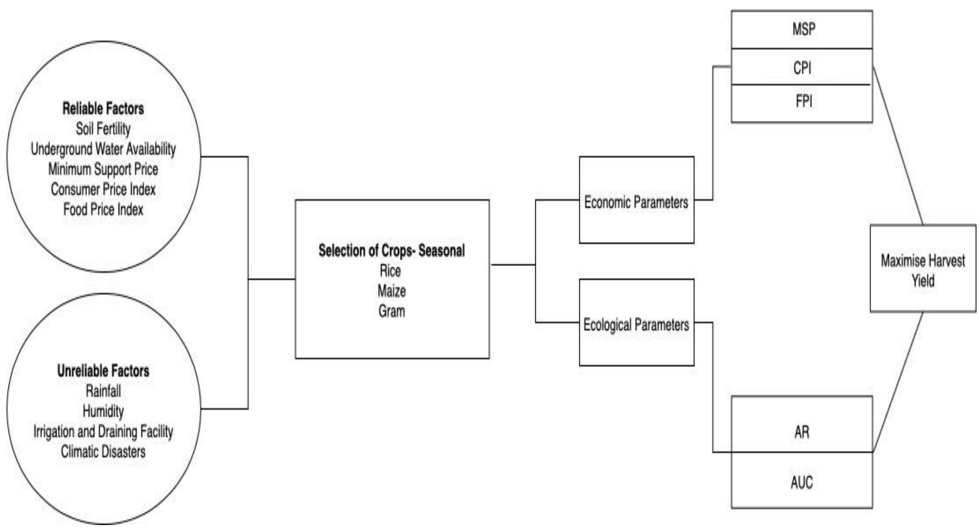

Figure 1 Factors affecting Crop Selection

In this work, a Crop Yield analysis model that facilitates farmers to take appropriate decisions considering the uncertainties of the factors involved. Annual rainfall, historical weather and price indices data are utilized to develop a semi- parametric yield analysis model for forecasting the Crop Yield.

Thus Multivariate Regression is used to analyse the influence of parameters (MSP, CPI, FPI, AR and AUC) on CY of Rice, Maize and Gram. Significant independent variables for each crop is identified and discussed.

Forecasting the CY of Rice, Maize and Gram is done using Multivariate Regression Model and the result is demonstrated using parity plot.

2 Literature Survey

Various methodologies and algorithms in Machine Learning are used to investigate and examine ecological spatio-transient information to plan and work with better models. The estimation procedure should be comprehended, overseen, and controlled. Various techniques are developed not only to depict and interpret the information in various perspectives but also to improve the effective management of resources.

Overall crop production can be increased by proper infrastructure for storing the grains and farm productivity, not only for catering the needs of growing population but also for exporting them globally [1]. 85.9 Million tons of Wheat is cultivated in the year 2011 and an 6.4% increase from the previous year. During 2011 itself cultivation in Rice depicts an 7% increase setting a new record of producing 95.3 million tons [2]. In India, agricultural produce worth of $39 billion was exported in 2013, making it the Sixth largest net exporter and Seventh largest agricultural exporter. 13.7% of nation's GDP was contributed by Agriculture, Forestry and Fishery. Numerous data mining techniques and its importance in the field of agriculture was surveyed in [3].

Ganguly et al. [4] used transient, spatial and spatiotemporal data mining techniques for mapping optimum weather conditions for better Crop Yield. The challenging factors considered were data for a long period, huge data size, non- linear dependency and uncontrollable behavior, determination of min and max threshold values. Thomas vandal et al [5] enumerated the effects of environmental changes in developing proper infrastructure due to ecological systems and power- plants. Contemporary Earth System Models (ESM) were kept running at spatial goals unreasonably coarse for evaluating the impacts in the limited area. Sirisha Adamala et al [6] enumerated the various difficulties encountered in gathering and dealing with 'Enormous Data' whose meaning changes from one territory to other.

Yuya Suzuki et al. [7] proposed a cloud support system for green house horticulture using support vector machine (SVM) based on agricultural irrigation system which adjusted quantity of water automatically. The nature of the soil and its moisture were predicted using sensors but the accuracy for the same was not mentioned. Ekasingh *et al.* [8] along with WEKA software used C4.5 model for building decision trees. The land unit, estimated production cost, land labour ratio and estimated profit as variables were considered to construct separate decision trees for dry and wet seasons.

The nature of the soil and its moisture were predicted using sensors but the accuracy for the same was not mentioned. Ekasingh *et al.* [9] along with WEKA software used C4.5 model for building decision trees.

Canonical correlation analysis on establishing relationship of various datasets and used to interpret those relationship in [10].Balkaya *et al.* [11] used Canonical correlation analysis for identifying the relationship between various plant characteristics and their yield components. Based on the relationship established, the plant characteristics were identified to determine yield per crop in winter season, but changes in climatic conditions were not discussed probabilistically.

The amount, regularity and amount of precipitation contrasts broadly between all the areas. Water bodies have defined drought based on the sources of water .Consequently, drought is not having single meaning Wilhite Glantz [12]. Hoyt [13] Characterized the dry season in muggy and semi-parched zones with a yearly precipitation inadequacy of 15%.

Palmer et al [14] characterized drought as a circumstance where the actual precipitation fell below the exact precipitation required for the current climatic conditions. [15],[16] characterized the outright dry season as the time for which the rain on any one day was less than 0.01inch for atleast 15 continuous days and the incomplete dry spell as the time of 29 sequential days and the mean precipitation of which did not surpass 0.01 inch per day.

3 Multivariate Regression

The main purpose of the research is to analyze the parameters influencing the Crop Yield. In reality, Crop Yield is influenced by a combination of ecological and economic factors. Multivariate Regression is exploited in the field of agriculture to find underlying associations that have been un-mined earlier.

The crops considered for this analysis are Rice, Maize and Gram. Agriculture is still the prime occupation in India hence exploiting the interrelationship among explanatory variables and the response variable. MSP, CPI, FPI, AUC and AR are the explanatory variables considered and the response variable Crop Yield. The data for a period of 30 years is collected from National Informatic Centre (NIC),Chennai and government web source [17][18][19][20][21] is used for the yield forecasting.

Multivariate Regression is used to analyze a response variable Z which changes with the variation in explanatory variables (Y_1, $Y_{2\ldots\ldots\ldots\ldots}Y_n$). This method ensures prediction of response variable from a given set of the explanatory variables. The Least-square fit, fits both linear as well as polynomial relationships and is used to predict Z for a set of unknown Y_1, Y_2 ... Y_n values. Multivariate Regression is used to compute the linearity and residual values to find the difference between the observed value of the dependent variable (Z) and predicted value ($Z^{\cap}$) of the dependent variable as given in Eqn. 1 and they are plotted as left over from the model fit.

$$\text{Residual}=\text{Observed Value}-\text{Predicted Value}=Z-Z^{\cap} \quad (1)$$

The square of the difference between the actual value and the predicted value of the dependent variable is said to be the residual sum of squares. The strength of a model is defined by R^2 Value and it portrays the model's strength. R^2 value ascertains the strength of interrelationship between the response variable and the explanatory variable. If the R^2 value is greater than 0.5, then the relationship between those variates is significant and nearing 1 defines perfect positive correlation.

Regression equation is explained as in Eqn.2 and weights are evaluated using cost function in MATLAB environment considering the parameters as dependent and independent as in Eqn.3

$$Z_i = \alpha + \delta_1 y_i(1) + \delta_2 y_i(2) + \ldots + \delta_n y_i(n) \quad (2)$$

Z_i= i^{th} component of the response variable y. n = number of explanatory variables. $y_i(j)$ = i^{th} component of j^{th} independent variable.

$$E(\alpha, \delta 1, \delta 2, \ldots, \delta n) = \frac{1}{2N}\sum_{i=1}^{n}(z - Z') \quad (3)$$

3.1 Multivariate Regression Applied to Rice Crop

Multivariate Regression is applied to rice crop data and the results are given in Heat map as in Fig 2 depicting the regression matrix in MATLAB environment between the variables and the Crop Yield of Rice. The R^2 values generally vary from -0.25 to 1. As the R^2 value increases from -0.25 to 1 the interrelationship between the variables also increases. The R^2 value of AR and AUC to Crop Yield is 0.91 and 0.78 respectively which is greater than 0.5 and which clearly indicates that the increase in AR and Cultivation area (AUC) proportionally increases the Crop Yield of Rice.

The R^2 value of MSP and CPI to Crop Yield is 0.87 and 0.95 respectively which is very closer to perfect positive correlation and shows MSP and CPI are highly significant in deciding the Crop Yield.

Figure 2 Regression matrix for Rice

As rice is a rain fed crop, equal spread of rain on the respective seasons is important. As AUC increases the production of Rice in tonnes increases which clearly indicates that increase in the cultivation area in optimum conditions is always preferred. The R^2 value of FPI to Crop Yield is 0.39 which is less significant.

3.2 Multivariate Regression Applied to Maize Crop

Based on the Multivariate Regression analysis, the regression matrix is constructed and shown in Fig 3. The R^2 value of AUC to Crop Yield is found to be 0.93 which clearly indicates that if the AUC increases then the Crop Yield of Maize also increases. The R^2 value of MSP and CPI to Crop Yield is found to be 0.92 and 0.97 which are very closer to 1 proving that the significance is very high in affecting the Crop Yield when compared to all other parameters. The R^2 value of FPI to Crop Yield is found to be 0.65 which is comparatively significant in affecting the Crop Yield. The R^2 value of AR is found to be -0.2 and is having the least influence portraying that Maize is not water fed crop.

The regression matrix for maize clearly shows that the AUC and Price Indexing factors influence much for the Maize Crop. If the MSP of the crop is increased by the government, the production of Maize is also found to increase.

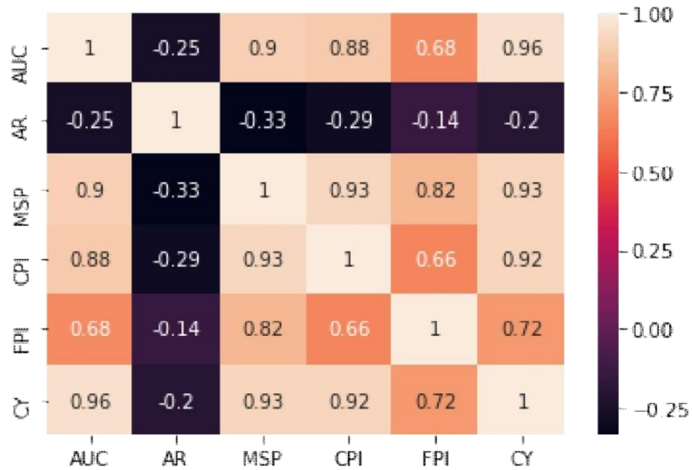

Figure 3 Regression Matrix for Maize

3.3 Multivariate Regression Applied to Gram Crop

The independent parameters AR, AUC, MSP, CPI and FPI are considered as explanatory variables and dependent parameter Crop Yield is considered as response variable. The results of Multivariate Regression are shown as regression matrix as given in Fig 4.

Figure 4 Regression Matrix for Gram

The R^2 value of AUC to Crop Yield is 0.84 which clearly indicates that increase in the cultivation area (AUC) will increase the Crop Yield of Gram. The R^2 value of MSP, CPI and FPI to Crop Yield is found to be 0.67, 0.63 and 0.68 respectively indicates the significance of economic parameters to the Crop Yield of Gram. The R^2 value of AR to Crop Yield is 0.11 has minimal influence depicting gram does not need more water for cultivation. Thus the AUC and economic factors significantly influences the Crop Yield of Gram.

4 Results and Discussions

The validity of the Multivariate Regression model and Prediction of CY are carried out using the agriculture data from 2007 to 2015. Multivariate Regression showcased that AR, AUC and MSP are the significant parameters influencing the CY of Rice. Similarly MVR shows the CY of Maize is significantly influenced by MSP, AUC and CPI and the CY of Gram is influenced by AUC and all three economic parameters (MSP,CPI and FPI). Since the Crop Yield is a function of the variables MSP, CPI, FPI, AR and AUC, its dependency is represented by Eqn.4 and Eqn.5.

Crop Yield = f(MSP,CPI,FPI, AR,AUC) (4)

$$CY = þ_0 + ((þ_1 *AUC) + (þ_2 *MSP) + (þ_3 *AR) + (þ_4 *FPI) + (þ_5 *CPI)) \quad (5)$$

4.1 Prediction Results of Rice

The prediction results of Multivariate Regression for rice are presented in Table 1, Table 2 and Figure.5 using MATLAB environment.

Table 1 Goodness of fit statistics (CY) - Rice

Regression of variable CY:	
Observations	32.000
Sum of weights	32.000
DF	26.000
R^2	0.974
Adjusted R^2	0.969
MSE	2.327
RMSE	0.515
MAPE	2.589
DW	2.456
Cp	6.000
AIC	64.392
SBC	73.186
PC	0.038

The prediction equation for forecasting the CY of rice is given in Eqn.6.

$$CY = -108.89+2.975*AUC+1.808*AR+6.352*MSP+0.265*CPI+0.133*FPI \quad (6)$$

In Table 1, the R^2 value is found to be 0.974 which shows that the model well represents the system in analysing and accurately predicting the agricultural attributes. The RMSE value is found to be 0.515 (Low RMSE value) indicates that the model predictions are very good. If AUC is increased by one unit, CY increases by 2.975 times. This denotes that, there is a positive effect of AUC in the prediction of crop yield along with the intercept value. If the annual rainfall rate is increased by one unit, the average Crop Yield increases by 1.808 times.

Table 2 Predictions and residuals (CY) - Rice

Observation	Weight	Actual (CY)	Pred(CY)	Residual	Std. residual
Obs1	1	53.630	52.126	1.504	0.598
Obs2	1	53.250	55.965	-2.715	-1.079
Obs3	1	47.120	48.646	-1.526	-0.607
Obs4	1	60.100	62.520	-2.420	-0.962
Obs5	1	58.340	60.643	-2.303	-0.916
Obs6	1	63.830	61.445	2.385	0.948
Obs7	1	60.560	61.993	-1.433	-0.570
Obs8	1	56.860	55.106	1.754	0.697
Obs9	1	70.490	69.535	0.955	0.380
Obs10	1	73.570	68.436	5.134	2.041
Obs11	1	74.290	76.722	-2.432	-0.967
Obs12	1	74.680	73.843	0.837	0.333
Obs13	1	72.860	71.148	1.712	0.681
Obs14	1	80.300	76.349	3.951	1.571
Obs15	1	81.810	79.653	2.157	0.858
Obs16	1	76.980	79.794	-2.814	-1.119
Obs17	1	81.740	81.681	0.059	0.024
Obs18	1	82.530	82.743	-0.213	-0.085
Obs19	1	86.080	88.013	-1.933	-0.768
Obs20	1	89.680	88.343	1.337	0.532
Obs21	1	84.980	86.917	-1.937	-0.770
Obs22	1	93.340	90.087	3.253	1.293
Obs23	1	71.820	77.528	-5.708	-2.269
Obs24	1	88.530	87.441	1.089	0.433
Obs25	1	83.130	83.759	-0.629	-0.250
Obs26	1	91.790	92.896	-1.106	-0.440
Obs27	1	93.360	93.891	-0.531	-0.211
Obs28	1	96.690	95.780	0.910	0.362
Obs29	1	99.182	101.547	-2.364	-0.940
Obs30	1	89.093	87.765	1.328	0.528
Obs31	1	95.325	94.900	0.425	0.169
Obs32	1	87.102	85.825	1.277	0.508

Since MSP is announced before the cropping season, Eqn.6 having high intercept value of 6.352 depicting its significance proving the fact that if the produce price is increased then the farmer tend to cultivate more of that crop. If the CPI is increased by one unit, the average crop yield increases by 0.265 times. The positive values indicate the direct relationship that exists between them. If the FPI is increased by one unit, the average crop yield increases by 0.133 times and it also indicates that the Crop Yield is not much dependent on the FPI. The mapping between actual CY and the predicted CY as shown in Figure.5

In Fig. 5, the data points are plotted to illustrate the closeness of the values obtained from the results analysed. Since, the data coordinates are more crowded near the linear curve it is therefore adequate to state that the predicted values and the actual values are almost the same.

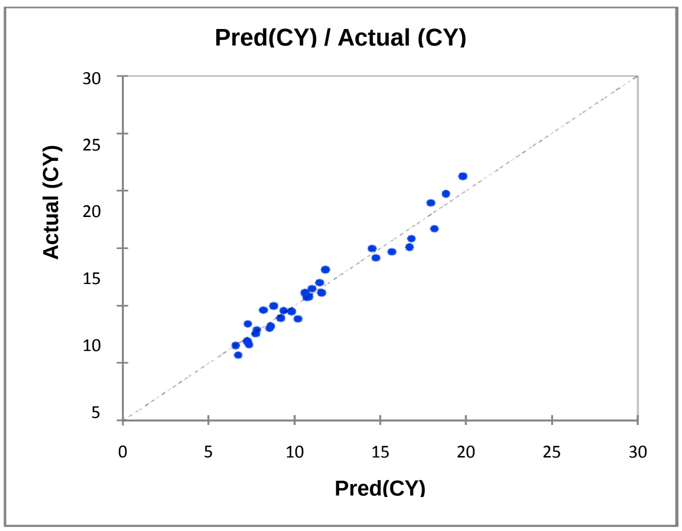

Figure 5 Parity plot - Rice

On scrutinizing the findings from the graphs and the factor equation, it is clear that predicted Crop Yield and the real Crop Yield are almost similar to one another and also proves that the factors considered for the research are appropriate. On evaluating the model fit, the R^2 value obtained is 0.974. This indicates that the model has the potential to explain 97.4% of variations pertinent to the training dataset.

4.2 Prediction Results of Maize

Similarly for Maize, Multivariate Regression is applied to predict the CY of Maize. The results of Multivariate Regression are given in Table 3, Table 4 and Figure.6. In Table 3, the RMSE value is found to be 0.939 (lowRMSE value) and indicates that the model predictions are very good.

Table 3. Goodness of fit statistics (CY) – Maize

Observations	32.000
Sum of weights	32.000
DF	26.000
R^2	0.956
Adjusted R^2	0.948
MSE	0.881
RMSE	0.939
MAPE	6.252
DW	2.322
Cp	6.000
AIC	1.313
SBC	10.108
PC	0.064

The Eqn. 7 shows the prediction equation for CY of Maize.

$$CY = -16.67+6.807*AUC+0.145*AR+9.898*MSP+3.107*CPI+1.899*FPI \quad (7)$$

The Eqn.7 shows the intercepts of the independent variables and its influence on estimating the CY of Maize. The intercept value of AUC and MSP are found to be 6.807 and 9.898 respectively. This shows that if 1 unit of AUC is increased the CY of Maize is found to increase by 6.807 times and similarly if MSP is increased by 1 unit then the CY of maize is found to increase by 9.898 times depicting that these 2 parameters are highly significant in influencing the CY of maize. The intercept value of AR is found to be 0.145 proving the fact that maize does not need much water for cultivation. The other two economic factors CPI and FPI also influence the CY of maize considerably having an intercept value of 3.107 and 1.899 respectively.

Table 4 Predictions and residuals (CY) – Maize

Observation	Weight	CY	Pred(CY)	Residual	Std. residual
Obs1	1	6.960	7.230	-0.270	-0.288
Obs2	1	6.900	7.271	-0.371	-0.395
Obs3	1	6.550	6.564	-0.014	-0.015
Obs4	1	7.920	7.797	0.123	0.131
Obs5	1	8.440	7.277	1.163	1.238
Obs6	1	6.640	7.352	-0.712	-0.759
Obs7	1	7.590	7.730	-0.140	-0.149
Obs8	1	5.720	6.704	-0.984	-1.049
Obs9	1	8.230	8.605	-0.375	-0.400
Obs10	1	9.650	8.177	1.473	1.569
Obs11	1	8.960	9.199	-0.239	-0.254
Obs12	1	8.060	8.540	-0.480	-0.511
Obs13	1	9.990	8.777	1.213	1.292
Obs14	1	9.600	9.351	0.249	0.265
Obs15	1	8.880	10.189	-1.309	-1.395
Obs16	1	9.530	9.819	-0.289	-0.308
Obs17	1	10.770	10.713	0.057	0.061
Obs18	1	10.820	10.845	-0.025	-0.027
Obs19	1	11.150	10.612	0.538	0.573
Obs20	1	11.510	11.012	0.498	0.530
Obs21	1	12.040	11.469	0.571	0.609
Obs22	1	13.160	11.789	1.371	1.461
Obs23	1	11.150	11.573	-0.423	-0.450
Obs24	1	14.980	14.508	0.472	0.503
Obs25	1	14.170	14.733	-0.563	-0.600
Obs26	1	14.710	15.675	-0.965	-1.028
Obs27	1	15.100	16.685	-1.585	-1.689
Obs28	1	18.960	17.950	1.010	1.076
Obs29	1	19.731	18.823	0.908	0.967
Obs30	1	16.720	18.149	-1.430	-1.523
Obs31	1	21.278	19.795	1.483	1.580
Obs32	1	15.855	16.809	-0.954	-1.016

In Fig. 6, the data points are sketched to illustrate the closeness of the values obtained from the analysis results.

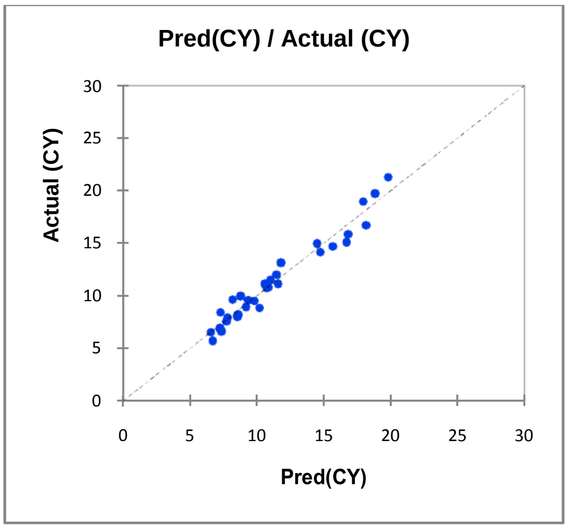

Figure 6 Parity plot – Maize

Since, the data coordinates are more crowded near the linear curve; it is therefore adequate to state that the predicted values and the actual values are almost the same. On scrutinizing the findings from the graphs and the factor equation, it is clear that predicted Crop Yield and the real Crop Yield are almost the same and also proves that the factors considered for the research are appropriate.On evaluating the model fit, the R^2 value obtained is 0.956. This indicates that the model has the potential to explain 95.6% of variations pertinent to the training dataset.

4.3 Prediction Results of Gram

Similarly for Gram, Multivariate Regression is applied to predict the CY of Gram. The results of Multivariate Regression are given in Table 5, Table 6 and Fig 7. In Table 5, the R^2 value is found to be 0.906 which shows that the model represents the system in analysing and accurately predicting the agricultural attributes. The RMSE value is found to be 0.363(low RMSE value) and indicates that the model predictions are very good.

Table 5 Goodness of fit statistics (CY) – Gram

Observations	31.000
Sum of weights	31.000
DF	25.000
R^2	0.906
Adjusted R^2	0.887
MSE	0.132
RMSE	0.363
MAPE	5.151
DW	2.678
Cp	6.000
AIC	-57.558
SBC	-48.954
PC	0.139

The Eqn.8 shows the prediction equation for CY of Gram.

$$CY = -3.472+7.907*AUC+0.536*AR+6.759*MSP+5.544*CPI+1.700*FPI \quad (8)$$

The Eqn.8 shows the intercepts of the independent variables and its influence on estimating the CY of Maize. If AUC is increased by one unit, CY increases by 7.907 times. This denotes that, there is a positive effect of AUC in the prediction of crop yield along with the intercept value. If the annual rainfall rate is increased by one unit, the average Crop Yield increases by 0.536 times. Since Gram needs less water for cultivation, the results also prove the same. Since MSP is announced before the cropping season, Eqn.8 having high intercept value of 6.759 depicts its high significance.
If the CPI is increased by one unit, the average Crop Yield increases by 5.544 times. The positive values indicate the direct relationship that exists between them. If the FPI is increased by one unit, the average crop yield increases by 1.700 times.

Table 6 Predictions and residuals (CY) – Gram

Observation	Weight	CY	Pred(CY)	Residual	Std. residual
Obs1	1	6.960	7.230	-0.270	-0.288
Obs2	1	6.900	7.271	-0.371	-0.395
Obs3	1	6.550	6.564	-0.014	-0.015
Obs4	1	7.920	7.797	0.123	0.131
Obs5	1	8.440	7.277	1.163	1.238
Obs6	1	6.640	7.352	-0.712	-0.759
Obs7	1	7.590	7.730	-0.140	-0.149
Obs8	1	5.720	6.704	-0.984	-1.049
Obs9	1	8.230	8.605	-0.375	-0.400
Obs10	1	9.650	8.177	1.473	1.569
Obs11	1	8.960	9.199	-0.239	-0.254
Obs12	1	8.060	8.540	-0.480	-0.511
Obs13	1	9.990	8.777	1.213	1.292
Obs14	1	9.600	9.351	0.249	0.265
Obs15	1	8.880	10.189	-1.309	-1.395
Obs16	1	9.530	9.819	-0.289	-0.308
Obs17	1	10.770	10.713	0.057	0.061
Obs18	1	10.820	10.845	-0.025	-0.027
Obs19	1	11.150	10.612	0.538	0.573
Obs20	1	11.510	11.012	0.498	0.530
Obs22	1	13.160	11.789	1.371	1.461
Obs23	1	11.150	11.573	-0.423	-0.450
Obs24	1	14.980	14.508	0.472	0.503
Obs25	1	14.170	14.733	-0.563	-0.600
Obs26	1	14.710	15.675	-0.965	-1.028
Obs27	1	15.100	16.685	-1.585	-1.689

Obs28	1	18.960	17.950	1.010	1.076
Obs29	1	19.731	18.823	0.908	0.967
Obs30	1	16.720	18.149	-1.430	-1.523
Obs31	1	21.278	19.795	1.483	1.580
Obs32	1	15.855	16.809	-0.954	-1.016

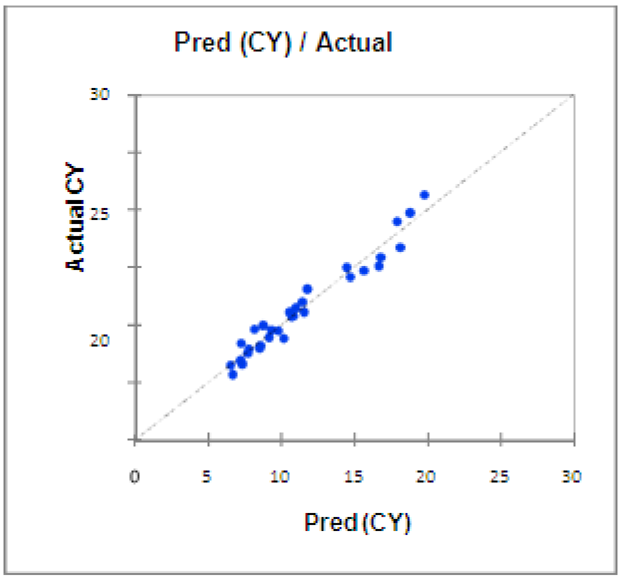

Figure 7 Parity plot – Gram

From the Fig. 7 , the data points are sketched to illustrate the closeness of the values obtained from the analysis results. Since, the data coordinates are more crowded near the linear curve; it is therefore adequate to state that the predicted values and the actual values are almost the same. On scrutinizing the findings from the graphs and the factor equation, it is clear that predicted Crop Yield and the real Crop Yield are almost the same and also proves that the factors considered for the research are appropriate. On evaluating the model fit, the R^2 value obtained is 0.906 which explains a very high correlation among the parameters considered.

5 Conclusion and Future Works

Thus multivariate regression model is used to exploit the influence of environmental parameters (MSP, CPI, FPI, AR and AUC) on CY of Rice, Maize and Gram. Significant independent variables for each crop is identified and discussed. Forecasting the CY of Rice, Maize and Gram is done using Multivariate Regression Model which depicts a model accuracy of 90.6% pertaining to training dataset. The agricultural data on other parameters like soil moisture, nature of the soil (clay/alluvial soil/acid/alkali/nutrient content),use of neem coated urea, use of hybrid seeds, proper timing for seeding, planting and harvesting ,water availability during the season, proper conditioning of the field etc., may be considered for better forecasting.

References

[1] J.S.Eales, B.K.Engel, R.J.Hauser,and S.R.Thompson, " Grain Price expectations of Illinois Farmers and Grain Merchandisers". American Journal of Agricultural Economics, vol.72 no.3, pp.701-708, 1990.
[2] S.Veenadhari, B.Misra, and C.D.Singh, "Machine learning approach for forecasting crop yield based on climatic parameters".In Computer Communication and Informatics (ICCCI),Coimbatore,pp. 1-5, 2014.
[3] V.Sellam, and E.Poovammal, "Prediction of crop yield using regression analysis" Indian Journal of Science and Technology, vol.9 no.38, pp.5, 2016.
[4] H.Patel, and D.Patel, "A brief survey of data mining techniques applied to agricultural data", International Journal of Computer Applications, vol. 95 no.9, 2014.
[5] Vatsavai, Ranga Raju,A.Olufemi Omitaomu, João Gama,V. Nitesh Chawla, and R. Auroop Ganguly " Knowledge Discovery from Sensor Data. Springer-Verlag Berlin Heidelberg ", 2010.
[6] T.Vandal, E.Kodra, S.Ganguly, A.Michaelis, R.Nemani, and A.R.Ganguly, "Deepsd: Generating high resolution climate change projections through single image super-resolution". In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1663-1672, 2017.
[7] S.Adamala, N.S.Raghuwanshi, A.Mishra, and R.Singh, "Generalized wavelet neural networks for evapotranspiration modeling in India". ISH Journal of Hydraulic Engineering, pp.1-13, 2017.
[8] Y.Suzuki, K.Nakamatsu, and H.Mineno, "A Proposal for an Agricultural Irrigation Control System Based on Support Vector Machine", 2013 IIAI

International Conference on Advanced Applied Informatics (IIAIAAI),IEEE, pp. 104-107, 2013.
[9] B.Ekasingh, K.Ngamsomsuke, R.A.Letcher, and J.Spate, “A data mining approach to simulating farmers' crop choices for integrated water resources management”, Journal of Environmental Management, vol. 77 no.4, pp.315-325, 2005.
[10] S.Klami, S.Virtanen, and S.Kaski, “Bayesian canonical correlation analysis”, Journal of Machine Learning Research, vol.14, pp.965-1003, 2013.
[11] A.Balkaya, S.Cankaya, and M.Ozbakir, “Use of canonical correlation analysis for determination of relationships between plant characters and yield components in winter squash (Cucurbita maxima Duch.) populations”. Bulgarian Journal of Agricultural Science, vol.17 no.5, pp.606-614, 2011.
[12] D.A.Wilhite, and M.H.Glantz, “Understanding: the drought phenomenon: the role of definitions”. Water International, vol.10 no.3, pp.111-120, 1985.
[13]Hoyt, Elizabeth Ellis "Consumption in our society" (1938).
[14] W.M.Alley, “The Palmer drought severity index: limitations and assumptions”. Journal of Climate and Applied Meteorology, vol.23 no.7, pp.1100-1109, 1984.
[15] P.H.Herbst, D.B.Bredenkamp, and H.M.G.Barker, “A technique for the evaluation of drought from rainfall data”,Journal of Hydrology, vol.4, pp.264-272, 1966.
[16]National Commission on Agriculture, “Report of the National Commission on Agriculture”, Ministry of Agriculture and Irrigation, 1976.
[17]https://data.gov.in/resources/all-india-area-weighted-monthly-seasonal-and-annual-rainfall-mm
[18]https://data.gov.in/resources/minimum-support-priceprocurement-price-crops-crop-year-basis
[19]https://data.gov.in/resources/annual-Food-price-index-agriculture-produce
[20]https://data.gov.in/resources/area-under-principal-crops-all-india-and-state-wise
[21]http://inflationdata.com/Inflation/Consumer_Price_Index/HistoricalCPI.aspx?reloaded=true

Biographies

V.Sellam has overall experience of 9+ years in the field of Information Technology,Science & Technology Projects and in Academic & Administration. She has been part of SRM group over 6 years and has served various academic and administrative roles. With six years of academic contribution and 4 years of research guidance and administrative contributions currently holds the position of Sr. Assistant Professor. She has obtained her Doctorate in Computer Science and Engineering from SRM Institute of Science and Technology, Chennai and Masters from Annamalai University. She has authored and co-authored research publications and books as her contributions. Her research alignment is broadly in the areas of Agricultural Data Mining, Speech Processing and Pattern Classification.

N.Kannan - Doctorate in Computer Science and Engineering having Twenty three years experience in teaching and research with good knowledge in the area of Computer science and Image Processing, currently working as Professor SRM Institute of Science and Technology, Chennai, India.

E.Poovammal-Doctorate in Computer Science and Engineering having Twenty five years experience in teaching and research with good knowledge in the area of Computer science and Data Mining, currently working as Professor SRM Institute of Science and Technology, Chennai, India.