Estimation of the Influence of Fertilizer Nutrients Consumption on the Wheat Crop yield in India- a Data mining Approach

A.Vjay Kumar, T.V. Rajini Kanth

ABSTRACT- The forecasting of agricultural yields is a challenging and desirable task for every nation. In Indian economy agriculture sector has a major role. In the total India rural population above seventy percent of the population depends on the agriculture to lead their lives. In the index of the Indian exports, agriculture exports stood at the fifth place. Today agriculture farmers are not only producing yields but also producing the agriculture data. This data can be collected, stored and analyzed for the useful information. In the present paper an attempt is made to apply the data mining techniques to extract useful information from the agriculture dataset of the annual measurements of the fertilizer nutrients consumed and wheat crop yields in India. The present experiment is based on the data collected from the sources like the Department of Agriculture and Statistics, Government of India and Department of Agriculture and Co operation, Government of India. The results of the present paper proved that the fertilizer nutrients consumed are the most influential factors of the wheat crop yield in India.

Keywords: Yield estimation, Data mining, precision agriculture, regression and regression analysis

I. INTRODUCTION

Data mining is an art and science of intellectual analysis of very large data sets for meaningful, previously unknown and potentially useful information. Today data mining is actively applied in the agriculture related areas. There is an emerging importance for the data mining techniques and methodologies in the field of "Data mining in Agriculture" since the effective and sustainable agriculture is the major issue in the recent years. It became essential for using the modern technologies such as GPS, GIS, remote sensing and data mining techniques to increase the agriculture production to meet the rapidly increasing population. Huge amounts of harvested data of the crops have to be analyzed and should be used to the maximum extent possible. Data mining allows extracting the most important information from such vast data to discover previously unknown and potentially useful patterns that may be relevant to current agricultural problems, thereby helping farmers to enhance the agriculture crops yields. There are a lot of opportunities for research on data mining in agriculture to develop practical applications for the enhancing of agriculture production.

Manuscript received December 2013.

A.Vijay Kumar, Department of Computer Science, University College of Sciences, Satavahana University, Karimnagar, Andhra Pradesh, India.

T.V.Rajini Kanth, Department of Computer Science and Engineering, Sree Nidhi College of Engineering and Technology, Jawahar Lal Nehru Technology, Hyderabad, Andra Pradesh, India.

Today data mining in agriculture is one of the very potential research areas. The current technologies are able to provide huge information on agricultural, and are analyzed to extract potentially useful information. The present research is aimed to apply data mining techniques on the dataset of wheat crop yield and fertilizer nutrients consumed to establish if any meaningful relationship can be found.

II. REVIEW OF LITERATUE

There is a wide variety of applications of data mining in agriculture like prediction of problematic wine fermentation, sorting of apples by water cores, optimizing pesticide usage and detection of diseases from the sounds issued by animals and many more.

M.Becker, K.H.Diekmann, J.K.Ladha, S.K.De Datta and J.C.G. Ottow in their paper [19] made an attempt to study the effect of NPK fertilizer nutrients consumed on the growth and nitrogen fixation of sesbania rostrata as a green manure for lowland rice. They also tried to know whether the integration of minerals and N, P and K fertilizer nutrients can significantly improve rice grain yield or not.

Suzuki.M, Kamekawa.k, Sekiya.S and Shiga.H in their paper [20] made an attempt to study the effect of continuous application of inorganic or organic fertilizers containing the same amount of N, P, K on rice soil, fertility and rice yields in the temperature region of central Japan. For their research work they were taken 60 years of agriculture yield data, fertilizer nutrients consumed and soil fertility data. In the research they identified that yields in the organic fertilizer plot was lower than the yields in the inorganic fertilizer plot for the initial 10 years. Later, yields of the organic fertilizer plot and inorganic fertilizer plot are same, sooner or after become higher after 30 years.

Georg Rub, Rudolf Kruse, Martin Schneider and Peter Wagner in the paper titled "Data mining with Neural Networks for Wheat Yield Prediction" made an attempt to predict the wheat yield by applying neural networks. The main objective of the research was to know whether it is possible to predict the wheat yield with neural networks or not. In the possible case they also thought to use the same logic to optimize the fertilizer usage in both the economic and environmental terms.

Georg Rub, Rudolf Kruse, Martin Schneider and Peter Wagner in the paper titled "Estimation of Neural Networks Parameters for Wheat Yield Prediction" made an attempt to estimate the parameters that influence the wheat yield. The feed-forward back propagation neural networks were evaluated to estimate the influential factors of wheat yield. Based on this prediction they made an attempt for economic and environmental optimization for example, fertilization etc.

III. OBJECTIVE OF STUDY

Estimating the influence of fertilizer nutrients (NPK) consumed on the yield of wheat crop in India.

IV. THE LIMITATIONS OF THE STUDY

1. The facts presented are based on the information collected through secondary sources and can't be generalized to all the agriculture output in the target area of the study.
2. The research can't depend upon the data mining only as it is one of the methods for the estimation of agriculture yield.

V. DATA SOURCES

Secondary sources include reports and documents from the Department of Agriculture and Statistics, India and the Directorate of Economics and Statistics, the Department of Agriculture and Co operation, India.

VI. THE EMPIRICAL STUDY ON THE ESTIMATION OF INFLUENTIAL FACTORS OF WHEAT CROP YIELD IN INDIA

Presently, the agriculture sector performance in our country has turned out to be quite dissatisfactory because of steep downfall in the growth rate of agricultural yield. The authorities are stressing for the needful measures to enhance the growth rate in agriculture, because of lower growth rate it results serious implications for large percent of India's population that depends upon agriculture for livelihood, and also affects growth rate of overall economy. High growth rate in non agriculture sector alone would not help India to realize a high growth rate in total economy. Therefore it is very much required to enhance growth rate in agriculture sector.

Agriculture growth depends upon identifying the factors, which are promoting the higher yield in agriculture. The factors promoting the growth in yield in agriculture are the proper use of inputs like fertilizers, the combination of Nitrogen (N), Phosphorus (P) and Potassium (k), also the conducive atmosphere with moderate temperature and irrigation facilities and rainfall. The factors underlying these are to be estimated with an appropriate tool so that an efficient mechanism is developed for promoting a higher agriculture yields. With the statistical analysis tools of data mining, we should identity the influential factors for the improvement in agriculture crop yield.

VII. THE ESTIMATION OF INFLUENCE OF THE FERTILIZER NUTRIENTS CONSUMED ON THE WHEAT CROP'S YIELD IN INDIA

To estimate the influential factors of wheat yield in India, a dataset of wheat is prepared by including the wheat yield and fertilizer nutrients (N, P, and K). To prepare the wheat dataset, yield and fertilizer nutrients consumed are collected from the Department of Agriculture and Statistics, India and the Department of Agriculture and Cooperation, India. By performing statistical analysis on this dataset, some interesting things about the wheat cultivation and its influential factors are identified. The following table shows the wheat yield and fertilizer nutrients consumed.

Fertilizer nutrients consumed and wheat yield

Year	N	P	K	Wheat-Yield
1950	58.7	6.9	-	663
1955	107.5	13	10.3	708
1960	210	53.1	29	851
1965	574.8	132.5	77.3	827
1970	1487	462	228	1307
1975	2148.6	466.8	278.3	1410
1980	3678.1	1213.6	623.9	1630
1985	5660.8	2005.2	808.1	2046
1986	5716	2078.9	850	1916
1987	5716.8	2187	880.5	2002
1988	7251	2720.7	1068.3	2244
1989	7386	3014.2	1168	2121
1990	7997.2	3221	1328	2281
1991	8046.3	3321.2	1360.5	2394
1992	8426.8	2843.8	883.9	2327
1993	8788.3	2669.3	908.4	2380
1994	9507.1	2931.7	1124.7	2559
1995	9822.8	2897.5	1155.8	2483
1996	10301.8	2976.8	1029.6	2679
1997	10901.8	3913.6	1372.5	2485
1998	11353.8	4112.2	1331.5	2590
1999	11592.7	4798.3	1678.7	2778
2000	10920.2	4214.6	1567.5	2708
2001	11310.2	4382.4	1667.1	2762
2002	10474.1	4018.8	1601.2	2610
2003	11077	4124.3	1597.9	2713
2004	11713.9	4623.8	2060.6	2602
2005	12723.3	5203.7	2413.5	2619
2006	13772.9	5543.3	2334.8	2708
2007	14419.1	5514.7	2636.3	2802
2008	15090.5	6506.2	3312.6	2907
2009	15580	7274	3632.4	2839
2010	16558.2	8049.7	3514.3	2988
2011	17300.3	7914.3	2525.5	3140

Table: w1

A chart view of the wheat yield and fertilizer nutrients consumed

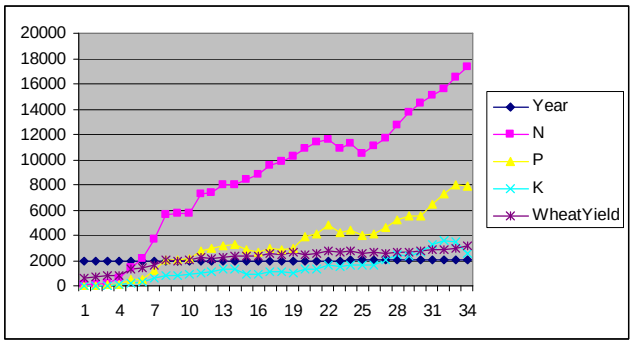

Chart: w1

Growth rates of yield of wheat during 1950-2011

The wheat production in India during the period 1950-2011 was almost continuously increased every year except two years, i.e., 1965-66 and 2005-06. During the years 1955-56, 1960-61, 1965-66, 1970-71, 1975-76, 1980-81, 1985-86, 1990-91, 1995-96, 2000-01,2005-06 and 2010-11 the growth rates are 6.79, 20.20, -2.82, 58.04, 7.88, 15.60, 25.52, 11.49, 8.86, 9.06, -3.29 and 14.09. There is a drastic increase in yield during the years 1970-71 and decrease during the years 2005-06 and 1965-66. The following table shows the wheat yield and growth rates of wheat yield.

Wheat yield and growth rates of wheat

Year	Yield of wheat	Growth rate of wheat
1950-51	**663**	**0.00**
1955-56	**708**	**6.79**
1960-61	**851**	**20.20**
1965-66	**827**	**-2.82**
1970-71	**1307**	**58.04**
1975-76	**1410**	**7.88**
1980-81	**1630**	**15.60**
1985-86	**2046**	**25.52**
1990-91	**2281**	**11.49**
1995-96	**2483**	**8.86**
2000-01	**2708**	**9.06**
2005-06	**2619**	**-3.29**
2010-11	**2988**	**14.09**

Table: W2

A Chart view of the wheat yield growth rates

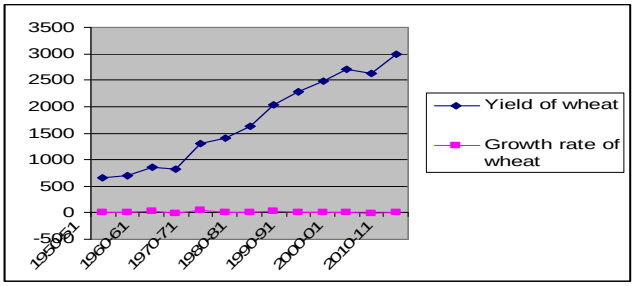

Chart: W2

Impact of fertilizer nutrients consumed on the yield of wheat in India

To estimate the impact of fertilizer nutrients consumed on yield of the wheat in India, a dataset of the wheat crop is constructed by including various factors like fertilizer nutrients, N, P, K and yield of the wheat. Using some statistical analysis tools a correlation and regression analysis is performed. The experiment revealed some interesting points about the impact of fertilizer consumption on the yield of wheat in India. The correlation between N, P, K and yield is very strong that indicates, when the fertilizer nutrients N, P, K consumption increases then yields of the wheat crop will also increases.

The results of the correlations of the fertilizer nutrients and yield

The correlation among the fertilizer nutrients, N, P, K and yield of the wheat crop are 0.96, 0.92 and 0.86 respectively. This indicates there is a high degree of correlation between fertilizer nutrients, N, P, K and yield of the wheat.

Year	Yield of wheat	Growth rate of wheat
1950-51	663	0.00
1955-56	708	6.79
1960-61	851	20.20
1965-66	827	-2.82
1970-71	1307	58.04
1975-76	1410	7.88
1980-81	1630	15.60
1985-86	2046	25.52
1990-91	2281	11.49
1995-96	2483	8.86
2000-01	2708	9.06
2005-06	2619	-3.29
2010-11	2988	14.09

Correlations

		WHEATYIE	N	P	K	YEAR
Pearson Correlation	WHEATYIE	1.000	.969**	.917**	.855**	.978**
	N	.969**	1.000	.975**	.921**	.963**
	P	.917**	.975**	1.000	.963**	.913**
	K	.855**	.921**	.963**	1.000	.872**
	YEAR	.978**	.963**	.913**	.872**	1.000
Sig. (2-tailed)	WHEATYIE	.	.000	.000	.000	.000
	N	.000	.	.000	.000	.000
	P	.000	.000	.	.000	.000
	K	.000	.000	.000	.	.000
	YEAR	.000	.000	.000	.000	.
N	WHEATYIE	34	34	34	33	34
	N	34	34	34	33	34
	P	34	34	34	33	34
	K	33	33	33	33	33
	YEAR	34	34	34	33	34

**. Correlation is significant at the 0.01 level (2-tailed).

Regression analysis

A regression analysis is performed on dependent variable wheat crop yield and independent variables, N, P and K (Fertilizer nutrients). Results of the regression analysis shows that yield of the wheat are strongly influenced by the fertilizer nutrients consumed during the study period. Results of regression analysis of fertilizer nutrients consumed and yield of the wheat crop are as follows.

Variables Entered/Removed [b]

Model	Variables Entered	Variables Removed	Method
1	P, K, N [a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: YIELD

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.971[a]	.943	.936	157.76

a. Predictors: (Constant), P, K, N

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regressio	L.0E+07	3	3422014	137.493	.000[a]
	Residual	22215.4	25	888.618		
	Total	L.1E+07	28			

a. Predictors: (Constant), P, K, N

b. Dependent Variable: YIELD

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	976.843	33.986		28.743	.000
	N	.119	.014	1.425	8.732	.000
	P	-6.42E-02	.043	-.349	-1.478	.150
	K	-5.01E-02	.058	-.121	-.869	.392

a. Dependent Variable: WHEATYIE

Clustering Using K-means

The k-means clustering technique is used to classify the wheat yield of India. K-means clustering with iterate and classify procedure classified the wheat crop yield data into two clusters with 30 cases in cluster one and 32 cases in cluster two. The initial cluster center points are 3140 and 653, the final cluster center points are became as 2473.73 and 1053 with the variation of 667.73 and 400.72 in the first and second cluster center points. The clustering output in the pie chart view is also shown in the figure below.

Initial Cluster Centers

	Cluster	
	1	2
WHEATYIE	3140.00	653.00

Iteration History[a]

Iteration	Change in Cluster Centers	
	1	2
1	596.185	468.371
2	47.401	44.553
3	22.680	23.099
4	.000	.000

a. Convergence achieved due to no or small distance change. The maximum distance by which any center has changed is .000. The current iteration is 4. The minimum distance between initial centers is 2487.000.

Final Cluster Centers

	Cluster	
	1	2
WHEATYIE	2473.73	1053.72

Number of Cases in each Cluster

Cluster	1	30.000
	2	32.000
Valid		62.000
Missing		.000

A pie Chart view of the wheat yield data

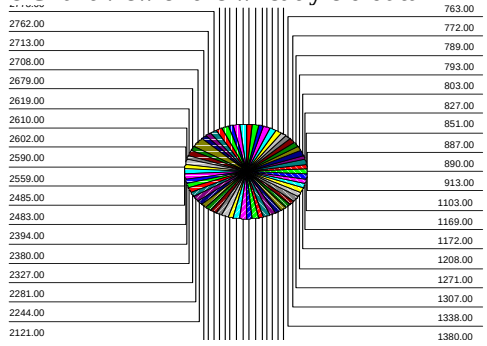

Chart: W4

CONCLUSION

The experiment conducted on the fertilizer nutrients consumed and the wheat yield data set proved that there is a strong correlation between them. During the years 1955-56, 1960-61, 1965-66, 1970-71, 1975-76, 1980-81, 1985-86, 1990-91, 1995-96, 2000-01,2005-06 and 2010-11, the growth rates were 6.79, 20.20, -2.82, 58.04, 7.88, 15.60, 25.52, 11.49, 8.86, 9.06, -3.29 and 14.09. There was a drastic increase in yield during the years 1970-71 and decrease during the years 2005-06 and 1965-66. In India, the wheat yield was also increased year to year except few years. Therefore, on the basis of results of the study we strongly believe and recommend that by increasing fertilizer nutrients consumption, yields of the wheat crop can be enhanced in India.

REFERENCES

[1] "Data mining Techniques for Predicting Crop Productivity – A review article" S.Veenadhari, Dr.Bharat Misra, Dr. CD Singh IJCST Vol. 2, Issue 1,March 2011.

[2] Chengquan Huang, Limin Yang, Bruce Wylie, and Collin Homer. "A strategy for estimating tree Canopy density using landsat 7 etm+ and high resolution images over large areas". In the proceedings of the Third International Conference on Geospatial Information in Agriculture and Forestry, 2001.

[3] Georg Ruß, Rudolf Kruse, Peter Wagner, and Martin Schneider. "Data mining with neural networks for wheat yield prediction. In Petra Perner, editor, Advances in Data Mining (Proc. ICDM 2008)", pages 47–56, Berlin, Heidelberg, July 2008, Springer Verlag.

[4] R S Deshpande "An Analysis Of The Results Of Crop Cutting Experiments,Agricultural Development and Rural Transformation Unit" Institute for Social and Economic Change February 2003.

[5] Georg Ruß "Data Mining of Agricultural Yield Data: A Comparison of Regression Models, ICDM'09",. Leipzig, Germany, July 2009.

[6] V. Ramesh and K. Ramr "Classification of agricultural land soils: A data mining approach", International Journal on Computer Science and Engineering (IJCSE) ISSN: 0975-3397 Vol. 3 No. 1 Jan 2011, 379.

[7] Anup K. Prasad a, Lim Chai b, Ramesh P. Singh a, b,*, Menas Kafatos b (2006). "Crop yield estimation model for Iowa using remote sensing and surface parameters". International Journal of Applied Earth Observation and Geoinformation 8 (2006) 26-33.

[8] B. Lalic, L. Pankovic, D. T. Mihailovic, M. Malesevic, I. Arsenic: "Crop models and its use in vegetation dynamic forecasting". In Proc. Of Institute of Field and Vegetable Crops, Vol. 44, pp. 317-323, 2007.

[9] B . Marinković, J. Crnobarac, D. Marinković, G. Jaćimović, D.V. Mircov, "Weather conditions in the function of optimal corn yield in Serbia and the Vojvodina province", in Proceeding of the 1st Scientific Agronomic Days, pp. 15-19, 2008.

[10] D. Pokrajac, T. Fiez, D. Obradovic, S. Kwek, Z. Obradovic, "Distribution comparison for site-specific regression modeling in agriculture", in Proc. 12th International Joint Conference on Neural Networks (IJCNN), pp. 3937-3941, 1999.

[11] Groten, S. M. E. (1993). "NDVI - crop monitoring and early yield assessment of Burkina Faso". International Journal of Remote Sensing, 14(8), 1495-1515.

[12] Hayes, M.J., Decker, W.L., 1996. "Using NOAA AVHRR data to estimate maize production in the United States" Corn Belt. Int. J.Remote Sense. 17, 3189-3200.

[13] Jorgensen, S.E., 1994. "Models as instruments for combination of ecological theory and environmental practice". Ecol. Model. 75-76,5-20.

[14] Mass, S. J., 1988. "Use of remotely-sensed information in agricultural crop growth models". Ecological Modeling. 41,247-268

[15] Moa,*, S. Liua, Z. Lina, Y. Xub, Y. Xianga, T.R. McVicarc (2005). "Prediction of crop yield, water consumption and water use efficiency with a SVAT-crop growth model using remotely sensed data on the North China Plain. Ecological Modelling" 183 301-322

[16] Monteith, J.L., 1972. "Solar radiation and productivity in tropical ecosystems". J.Appl. Ecol. 9,747-766.

[17] Monteith, J.L., 1977. "Climate and the efficiency of crop production Britain". Phil. Trans.Roy.Soc.Lond. B 281,277-294.

[18] Reynolds, M. Yittayew, D. C(2000).Slack, “Estimation of crop yields and production by integrating the FAO Crop Speci. C Water Balance model with real-time satellite data and ground-based ancillary data”. Int.g. remote sensing, 2000, vol,21, no.18,3487-3508

[19] M. Becker, K. H. Diekmann, J. K. Ladha, S. K. De Datta, J. C. G. Ottow, “Effect of NPK on growth and nitrogen fixation of Sesbania rostrata as a green manure for lowland rice (Oryza sativa L.)” Plant and Soil, April 1991, Volume 132, Issue 1, pp 149-158

[20] Suzuki.M, Kamekawa.k, Sekiya.S and Shiga.H in the paper titled “Effect of continuous application of organic or inorganic fertilizer for sixty years on soil fertility and rice yield in paddy field” Trans.14th Int.congr. soil sci. 4:14-19, Vol.4

Mr.Ananthoju Vijay Kumar perusing Ph.D from Acharya Nagarjuna University, Guntur, Andhra Pradesh since 2008 to till date. He has obtained his Master of Computer Application degree from Osmania University, in June 2006 and Bachelor of Computer Application from Kakatiya University, in May 2003. His specialization area in research is Data mining in Agriculture. He is presently working as a Contract Lecturer in the Department of Computer Science, University College of Sciences, Satavahana Univeristy, Karimangar, Andhra Pradesh, India from August 2007 to till date. Prior to that he worked as a Assistant Professor, in the department of Computer Applications, SRM PG college, Karimnagar, Andhra Pradesh, India, from June, 2006 to August 2007. His total teaching experience is 8 years. He had attended several national and International seminars and conferences for paper presentation. He also published several articles in the various journals.

Dr. T. V. Rajini Kanth has obtained his Ph.D. degree in C.S.E. branch from Osmania university, Hyderabad in July, 2008 and M. Tech. (C.S.E.) degree from Osmania University, Hyderabad in January, 2001. His specialization area in research is “Spatial Data mining”. He obtained his PGDCS degree from HCU, Hyderabad in 1996. He received his M. Sc. (Applied mathematics) degree in the year 1989 from S.V. University, Tirupati as University Ranker. He is working as a Professor and Head, Department of IT, GRIET, Hyderabad since 2008 June. His writings have appeared in numerous Professional conferences and Journals (International journals-21, national level-4,). He was an author for few books i.e Artificial Intelligence etc. He has conducted two International conferences namely ICACT-08, and ICACM-11 at GRIET, Hyderabad and also acted as session chair for many conferences. He was convener for Pragnya-08, a national level technical fest at GRIET. Apart from that many National level ROBO workshops were conducted namely RoboTrix, eTrix, iTrix, LogiTrix, VisionTrix and Haptic Robotic arm in association with IIT Bombay and Technophilia etc.. Apart from these presently guiding around 22 Ph.D. scholars at various universities namely, JNTUH, JNTUK, JNTUA and NU. He is currently editor for two journals. He was called for about 50 AICTE sponsored workshops as resource person.