

Review- Role of Data Mining in Agriculture

Raorane A.A.¹ Kulkarni R.V.²

¹*Department of computer science,
Vivekanand College,
Tarabai park Kolhapur INDIA.*

²*Head of the Department,
Chh. Shahu Institute of business Education and Research Centre
Kolhapur. 416006 INDIA*

Abstract -Reliable and timely forecasts of crop production are required for various policy decisions relating to storage, distribution, pricing, marketing, import-export, etc. The official forecasts(advance estimates) of major cereal and commercial crops are issued by Directorate of Economics and Statistics. However, these advance estimates are only gustomates and not the objective estimates. In working out these estimates there is lots of subjective assessment based on several qualitative factors. There is thus a need to develop statistically sound objective forecasts of crop acreage and production.

The advances in computing and information storage have provided vast a most of data. The challenge has been to extract knowledge from this raw data; this has lead to new methods and techniques such as data mining that can bridge the knowledge of the data to the crop yield estimation. This research aimed to assess these new data mining techniques and apply them to the various variables consisting in the database to establish if meaningful relationships can be found.

Keywords- Yield estimation, Data mining, regression analysis, crop cutting experiments

INTRODUCTION :-

Agriculture in India has a significant history. Today, India ranks second worldwide in farm output. Agriculture and allied sectors like forestry and fisheries accounted for 16.6% of the GDP 2009, about 50% of the total workforce. The economic contribution of agriculture to India's GDP is steadily declining with the country's broad-based economic growth. Still, agriculture is demographically the broadest economic sector and plays a significant role in the overall socio-economic fabric of India.

Indian agriculture is known for its diversity which is mainly result of variation in resource and climate, to topography and historical, institutional and socio economic factors. Policies followed in the country and nature of technology that became available over time has reinforced some of the variations resulting from natural factors. As a consequence, production performance of agriculture sector has followed on uneven path and large gaps have development in productivity between different geographic locations across the country.

Agriculture as a business is unique crop production is dependent on many climatic, geographical, biological political and economic factors that are mostly independent of one another. This multiple factor introduces risk. The efficient management of these risks is imperative for the successful agricultural and consistent output of food.

The Agricultural yield is primarily depends on weather conditions, diseases and pests, planning of harvest

operation. Effective management of these factors is necessary to estimate the probability of such unfavorable situation & to minimize the consequences. Accurate and reliable information about historical crop yield is thus vital for decisions relating to agricultural risk management.

Historical crop yield information is also important for supply chain operation of companies engaged in industries that use agricultural produce as raw material. Livestock, food, animal feed, chemical, poultry, fertilizer pesticides, seed, paper and many other industries use agricultural products as intergradient in their production processes. An accurate estimate of crop size and risk helps these companies in planning supply chain decision like production scheduling. Business such as seed, fertilizer, agrochemical and agricultural machinery industries plan production and marketing activities based on crop production estimates.[1],[13]

APPLICATION

In past decades, IT has become more & more part of our everyday lives. With IT improvements in efficiency can be made in almost any part of industry and services, now days this is especially true for agriculture. A farmer now days harvest not only crops but also growing amounts of data. These data are precise & small in scale.

However, collecting large amounts of data often is both a blessing and a curse. There is a lot of data available containing information about certain asset. Here soil and yield properties, which should be used to the farmers advantage. This is a common problem for which the term data mining has been coined. Data mining techniques aim at finding those patterns or information in the data that are both valuable and interesting to the farmer.

A common specific problem that occurs is yield prediction. As early into the growing season as possible, a farmer interested in knowing how much yield he is about to expect. In the past, this yield prediction has actually relied on farmer's long-term experience for specific yield, crops and climatic conditions. However, this knowledge might also be available, but hidden in the small-scale. Precise data which can now days collected in seasons using a multitude of seasons.

Upgrading and stabilizing the agricultural production at a faster pace is one of the basic conditions for agricultural development. Productions of any crop lead either by attention of area or improvement in productivity or both. In India, the possibility of extending the area under any crop, almost, does not exist except by restoring to increased cropping intensity or crop substitution. Moreover, area and

productivity of different crops are the results, and as well as the reflection of the combined effect of many factors like agro-climatic conditions resource endowment technology level, techniques adopted infrastructure, social & economic conditions many schemes have been devised to maximize the productivity of various crops in different agro-climate region, state departments, credit institution, seed/fertilizer pesticide agencies & many other partners in public & private sections are actively engaged in enhancing the productivity of different crops in different regions and under different condition. However fluctuations in crop productivity continue to dog the sector and create severe distress.

Estimation of productivity of different crops is one of the important activities undertaken by the government departments in order to monitor the progress of the sector & provide insurance to the sector. Revenue, agriculture & Economics & statistics departments are jointly involved in the estimation process. Researcher & many other agencies use the data so generated by the Government departments. But these are usually available only in an aggregate form & maximum of taluka level satellite images of crop slate are being used increasingly to estimate the area but productivity data have to come from crop cutting experiment.

Article 243-9 of constitution of India requires the panchayat Raj institutions to be the decision making bodies in various aspects of agricultural sector and especially the implementation of the schemes. Crop Insurance is one of the important schemes of the agricultural sector. The debate in implementation of this scheme indicated requirement of the yield estimates of lower than the taluka level and especially of panchayat level. [3]

REVIEW OF LITERATURE:-

For examining the null hypothesis and its importance in crop insurance, an extensive literature survey was conducted. Research papers in journals like American Journal of Agricultural Economics, Canadian Journal of Agricultural Economics, Agribusiness, and North Central Journal of Agricultural Economics were reviewed to analyze the findings of various researchers working in the general area of frequency distributions for historical crop yields.

In addition to the literature survey, a quantitative analysis of actual yields is considered necessary to test the null hypothesis. As part of this testing, data on historical a crop yields for sugarcane and soybeans were collected from India from Kolhapur districts for analysis. This data set used for testing the assumption of normality crops were selected for data collection and analysis because of their economic importance to Kolhapur district.

Day (1965) – He suggest that yield distribution in agricultural crop do not exhibit normality. [5]

Dorfan (1991) Argues that a large amount of agricultural economic data is inconsistent with the assumption of normality of Crop yield distribution. [6]

The importance of skew ness in crop yield forecasting is specified by Gallagher by Suggesting that ignoring skewed distribution will lead to underestimation of most likely yields.[7]

Just and Weinenger (1991) have studied country level data and based on their experiment the disagreed with crop yield distributions are non normal argued that the evidence available to date is not enough to disapprove normality of crop yields. [9]

Norwood B. Roberts, Lusk (2004) – By their studies it was observe that the semi parametric model ranked highest for forecasting purposes.[8]

Ramirez, Misra & Field (2003) – Studied yield distribution of these crops and conclude that they are non normal and left skewed. [10]

From the research article “Data mining of agricultural yield Data: A comparison of regression models” George RuB express that large amount of data which is collected and stored for analysis. Making appropriate use of these data often leads to considerable gains in efficiency and therefore economic advantage. This paper deals with appropriate regression techniques on selected agriculture data.[11]

“Classification of agricultural land soils: A data mining approach” In this research paper V. Ramesh and K. Ram explains comparison of different classifiers and the outcome of this research could improve the management and systems of soil uses throughout a large fields that include agriculture, horticulture, environmental and land use management.[12]

D. R. Mehata and others are worked on “Rainfall variability analysis and its impact on crop productivity” In this case study they collected the weekly rainfall data and number of rainy days recorded at the main Dry farming research station from 1958 to 1996 (39 yrs). The correlation and regression studies were worked out using rainfall(x) as independent variable and yield(y) as dependent variable to derive information on rainfall-yield relationship and to develop yield prediction model for important crops.

From “Generalized software tools for crop area estimation and yield forecast” Roberto Benedetti and others describes the procedure that leads to the estimates of the variables of interest, such as land use and crop yield and other sampling standard deviation, is rather tedious and complex, till to make necessary for statistian to have a stable and generalized computational system available. The SAS is also often the ideal instrument to face with these needs, because it permits the handling of data effectively and provides all necessary functions to manage easily surveys with thousands of micro data. This paper focus on the use of this system in different steps of the survey: sample design, data editing and estimation. The information produced is however, available for one user only, the manager of the survey.[13]

“Risk in Agriculture: A study of crop yield distribution and crop insurance” by Narsi Reddy Gayam in his research study examines the assumption of normality of crop yields using data collected from INDIA involving sugarcane and Soybean. The null hypothesis (Crop yield are normally distributed) was tested using the Lilliefore method combined with intensive qualitative analysis of the data. Result show that in all cases considered in this thesis, crop yield are not normally distributed.[13]

CONCLUSION

In view of this, there is a need for an objective methodology for pre-harvest crop forecasting. This involves building up suitable forecast model(s) which has certain merits over the traditional forecasting method. These merits include the objectivity of the forecast and its ability to provide a measure of reliability which a traditional forecast method cannot provide. This, as such, calls for the necessity of objective methods for pre-harvest forecast of crop yields 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] Chapman P. Gleason LARGE AREA YIELD ESTIMATION/FORECASTING USING PLANT PROCESS MODELS By Chapman P. Gleason For Presentation at the 1982 Winter Meeting AMERICAN SOCIETY OF AGRICULTURAL ENGINEERS Palmer House, Chicago, Illinois December 14-17, 1982
- [3] 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
- 4) Ramesh Chand, Sanjeev Garg and Lalmani Pandey “Regional Variations in Agricultural Productivity A District Level Study” in 2009 for National Professor Project
- 5) Day, R. H., “Probability distribution of field crop yields” Journal of farm Economics 47(1965) 7B – 741.
- 6) Dorfman J.H. “should normality be a normal assumption?” Economic letters 42 (1993) 143 – 147
- 7) Gallagher P. “U.S. Soyabean yields: estimation and forecasting with nonsymetric disturbance” American Journal of Agriculture Economics 69. (Nov. 1987) : 796 .803.
- 8) Norwood B. Roberts, M.C. “Lusk J. L. “ Ranking Crop yield model using out of sample likely wood functions”. American Journal of Agriculture Economics 86. (4) (Nov. 2004)1032.1043
- 9) Just R. E. Weinenger Q. “Are crop yields normally distributed” American Journal of Agriculture Economics 81. (May 1999): 287. 304.
- 10) Ramirez, Misra & Filed – “ Crop yield distribution revisited” American Journal of Agriculture Economics 2003. Volume 85: 108
- 11) Georg Ruß Data Mining of Agricultural Yield Data: A Comparison of Regression Models, ICDM’09., Leipzig, Germany, July 2009
- 12) 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
- 13) Rainfall variability analysis and its impact on crop productivity Indian agriculture research journal 2002 29,33.,8) SPRS Archives XXXVI-8/W48 Workshop proceedings: Remote sensing support to crop yield forecast and area estimates GENERALIZED SOFTWARE TOOLS FOR CROP AREA ESTIMATES AND YIELD FORECAST by Roberto Benedetti A, Remo Catenaro A, Federica Piersimoni B
- 14) “Risk in Agriculture: A study of crop yield distribution and crop insurance” by Narsi Reddy Gayam Thesis (M. Eng. in Logistics)--Massachusetts Institute of Technology, Engineering Systems Division, 2006. Includes bibliographical references (leaves 52-53).
- 15) Gazetteer of Kolhapur District (2001)
- 16) Aditya, Kaustav (2008). Forecasting of crop yield using discriminant function technique. M.Sc. thesis, PG School, IARI, New Delhi.
- 17) Agrawal, Ranjana, Jain, R.C. and Singh, D.(1980). Forecasting of rice yield using climatic variables. Ind. J. Agric. Sci., 50 (9), 680-684.
- 18) Agrawal, Ranjana and Jain, R.C. (1982). Composite model for forecasting rice yield. Ind. J. Agric.Sci., 52 (3), 189-194.
- 19) Agrawal, Ranjana, Jain, R.C. and Jha, M.P. (1983). Joint effects of weather variables on rice yields. Mausam, 34 (2), 177-181.
- 20) Agrawal, Ranjana, Jain, R.C. and Jha, M.P. (1986). Models for studying rice crop weather relationship. Mausam, 37 (1), 67-70.