International Journal of Engineering and Innovative Technology (IJEIT)
Volume 2, Issue 2, August 2012

Identification of Agricultural Production Areas in Andhra Pradesh

J. Rajendra Prasad, P. Ravi Prakash, S. Sai Kumar, M. Sundara Babu K. Swarupa Rani

Abstract*—This paper will propose a data mining framework for the identification of agricultural production areas in Andhra Pradesh. The main aim of this paper indicates the analysis of agricultural datasets compared to currently using statistical methods. The online analytical process is used to take horizontal axis and data, information, knowledge and wisdom are used as a vertical axis. In this paper we are describing the DM Framework development, description, components and experimental results. This paper is specially used for crop prediction; planting strategist test results are very much useful to the formers to understand market needs and planting strategies.

Keywords: **Data Mining, Framework, Statistics, OLAP.**

I. INTRODUCTION

Data mining has attracted a great attention in the information industry and in society as a whole in recent years, due to wide availability of huge amount of data and the imminent need for turning such data into useful information and knowledge. The information and knowledge gained can be used for application ranging from market analysis, fraud detection, to production control, disaster management and science exploration. Data mining (DM) is an automated prediction and analytical process which is involved in the transformation of data into useful information. This is achieved through uncovering latent patterns which are hidden in enormous amounts of related data available in various databases and data warehouses.[1] The artifacts which are produced from the data mining process may then be utilized either in automated decision support mechanisms or assessed manually by decision makers [2]. The exponential growth of available data has been made possible by an ever increasing capability for data collection and storage capacity in computer hardware Advancements in technology. This development has occurred over the last two decades and is due to the availability, affordability and effectiveness of computer storage devices in particular, the hard-drives and the associated read-write access times [3]. These advances in data availability through facilitated storage, data availability and data diversity have resulted in an increase in the complexity and interrelationships of the data entities. Consequently, traditional user-driven analysis of the storehouses of data by statisticians has become increasingly difficult, creating a demand for an automation of the process [3]. Furthermore, the latent patterns hidden within the data may not be uncovered through simple statistical means [4]. Data mining solves these problems by providing a richer set of tools capable of discovering the patterns and inferring new knowledge. This fact has seen the proliferation of data mining as an emergent pattern matching and prediction tool [5]. Many agricultural based research organizations have initiated programs which no longer depend solely on statistical analysis for such recommendations but have incorporated data mining as a feasible alternative. For example the Agricultural Ministry of Pakistan have made the successful transition already and have reported the benefits of increased prediction accuracy as part of their crop management strategy [6]. There have also been other research studies which have sought to include OLAP to enhance the data mining experience [7]. There have been instances of generalization based data mining where object cube models and OLAP are investigated [8]. A more recent study where OLAP is used for quick analysis of aggregates of agricultural data was done in Pakistan in 2009 in the field of pest management in cotton crops [9]. This study utilizes a DM framework which may be used to enhance the interrogation of agricultural data and make recommendations for it usage in an agricultural context.

II. DATA MINING FRAMEWORK

In order to understand what a framework is, it is necessary to understand that system data do not exist in isolation but may be related to other data in a variety of ways due to a them sharing common features [12]. Although the data from different systems may have common features, they appear to be outwardly unrelated or related in uncharacteristic and undescribed ways [13]. Consequently a framework is a well structured and re-definable specification that permits the identification of t he common properties whereby the meanings for the common concepts can be identified and understood by creating models from common abstractions. Frameworks may also be defined in terms of interfaces and human computer interaction (HCI) where systems and or components combine with respect to interpretations of the spatial, relational and constructive domains [15]. In the spatial approach, the position and orientation of the building blocks of the framework are significant in terms of the interface parameters. The relational approach deals with logical and abstract considerations. Frameworks confirm more visibly to the constructive approach where modular elements arc assembled and connected together. In addition, frameworks must be re-usable [16]. In order for them to achieve this property within a context, frameworks must therefore incorporate levels of abstraction, granularity and specificity as depicted in figure 1 [17].

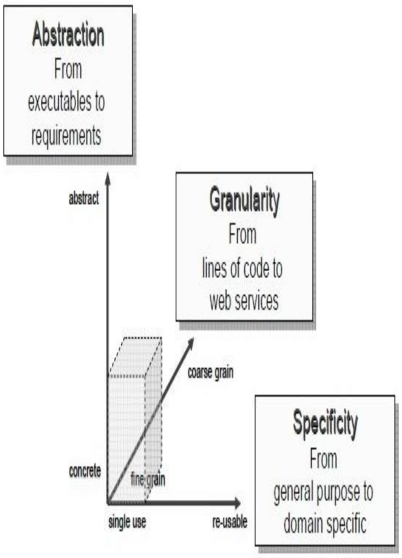


Fig 1.Aspects of Framework & Software Design [17].

Frameworks have been used in many previous studies in a host of different research areas such as software [18], including many in the field of data mining such as the theoretical framework used for pattern recognition [19], a framework for defining classification rules in a network intrusion detection system (IDS) context [20], the exploration of the parameter of 'transversal endurance' through the use of a classification hierarchy as a framework [21] and geographic measurement frameworks [22].

III. Development of the DM framework

The DM framework was developed after previous studies concluded that data mining had distinct advantages over single statistical methods of analyzing data where the hypothesis testing paradigm was the norm [26]. The idea to include online analytical processing (OLAP) as part of the framework arose from other previous studies in the agricultural data mining area which have combined data mining to other data analytical processes [6]. Metaphorically speaking, the use of DM techniques inspired the creation of the DM framework from a hypothesis generation perspective especially in terms of exploratory data mining (EDM) [27]. In other words, if a bottom-up approach is equated to the hypothesis generation perspective and a top-down approach is equated to a hypothesis testing perspective, then the DM framework occupies a pathway that is a mixture of both perspectives. In this regard, the DM framework conforms to the defining characteristic of abstraction where the concrete aspect is represented by case studies (data) and the abstract is the actual DM framework itself together with the insights (information, knowledge and recommendations) that the use of the framework provides.

The constructed framework is based on the data, information, knowledge and wisdom continuum [28].The understanding of this continuum is depicted graphically as a natural progression [29] as described in Figure 2

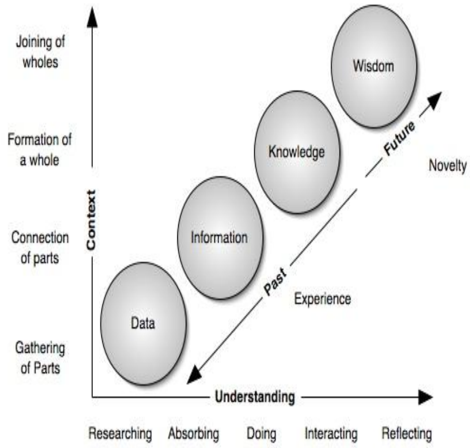


Fig 2. DM Framework Descriptions

The DM framework assumes a logical process of data capture, storage, processing and customized reporting to end-users. This logical progression assumes top-down significance in relation to its construction. The volume and content of the data interrogated through the use of data mining tools is used with the aim of extracting those components of the data that would be considered best-practices. Best practices is a term that may be used to describe the 'nuggets' of information that could be applied to various and differing agricultural settings while still remaining essentially true.

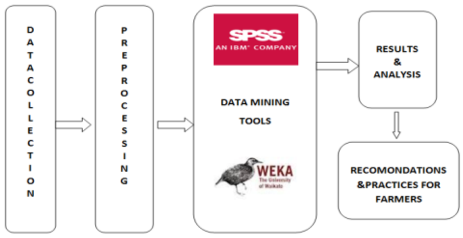


Fig 3.The Proposed Data Mining Framework

IV. THE DM FRAMEWORK COMPONENTS

The methodology used in developing this framework is constituted of several related components. The exploration of the associations and interactions with and between each of the components has enabled a framework to be developed. The components are made of the processes of Data collection, Data Preprocessing & Storage, Data Mining Tools, Results& Analysis, and Practices.

Data Collection

The data that will be used as part of case studies for the validation of the DM framework has been captured over various agricultural sites [30].

Data Preprocessing: It is done to improve the efficiency and ease of mining process. It contains four steps.

1. Data Cleaning: Removes noise and corrects inconsistencies.
2. Data Integration: Merges data from multiple sources into a coherent data store like a data warehouse/data cube.
3. Data Transformation: Improves the accuracy and efficiency of mining algorithms
4. Data Reduction: Reduces the data size by aggregating, eliminating redundancy, or clustering.

Data Storage

The variously related data elements have been stored as separate datasets in a common database and made available to the DM software tools for importing into simple data formats such as excel, csv, arff

Data Mining Tools

Data Mining (DM) tool is part of the DM framework that make up the active transformations of the data into views. Both data mining the tools of presentation of the data into useful information. The two mechanisms offer contrasting views of the data. Data mining allows a microscopic Data mining has the tendency towards an inductive approach to problem solving. Validation of the framework may be made through the use of such DM tools as Weka and Clementine where hidden trend and latent trends are sought. Here we used Clementine tool for performing results.

Results and Analysis

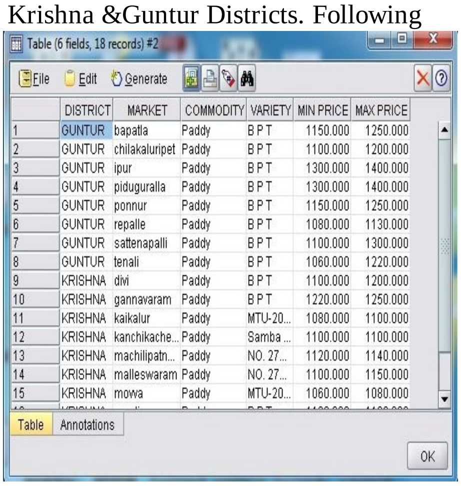
Calculate the results and identify the market requirements and identify the production areas and frequent item sets by using the Aprioi algorithm. And suggest the farmers to cultivate according to customer needs.

Recommendations to the Formers

Based on results and analysis suggest the Useful information to the farmers to benefit more results from latest updates etc.

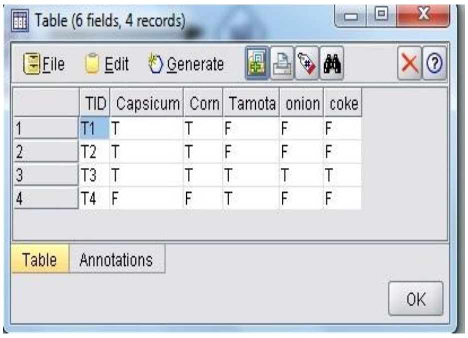
V. EXPERIMENTAL RESULTS

The following Figure4 depicts the production areas of Paddy in Krishna &Guntur Districts. Following



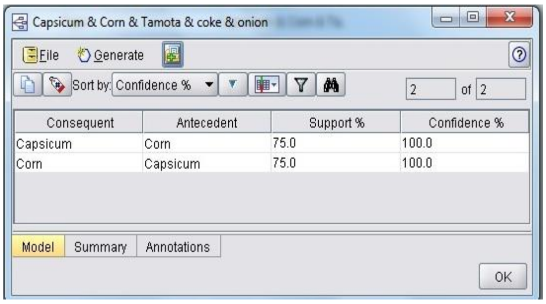
	DISTRICT	MARKET	COMMODITY	VARIETY	MIN PRICE	MAX PRICE
1	GUNTUR	bapatla	Paddy	B P T	1150.000	1250.000
2	GUNTUR	chilakaluripet	Paddy	B P T	1100.000	1200.000
3	GUNTUR	ipur	Paddy	B P T	1300.000	1400.000
4	GUNTUR	piduguralla	Paddy	B P T	1300.000	1400.000
5	GUNTUR	ponnur	Paddy	B P T	1150.000	1250.000
6	GUNTUR	repalle	Paddy	B P T	1080.000	1130.000
7	GUNTUR	sattenapalli	Paddy	B P T	1100.000	1300.000
8	GUNTUR	tenali	Paddy	B P T	1060.000	1220.000
9	KRISHNA	divi	Paddy	B P T	1100.000	1200.000
10	KRISHNA	gannavaram	Paddy	B P T	1220.000	1250.000
11	KRISHNA	kaikalur	Paddy	MTU-20...	1080.000	1100.000
12	KRISHNA	kanchikache...	Paddy	Samba ...	1100.000	1100.000
13	KRISHNA	machilipatn...	Paddy	NO. 27...	1120.000	1140.000
14	KRISHNA	malleswaram	Paddy	NO. 27...	1100.000	1150.000
15	KRISHNA	mowa	Paddy	MTU-20...	1060.000	1080.000

Fig 4 Present Market Needs Based On Aprioi Algorithm On Sample Data.



	TID	Capsicum	Corn	Tamota	onion	coke
1	T1	T	T	F	F	F
2	T2	T	T	F	F	F
3	T3	T	T	T	T	T
4	T4	F	F	T	F	F

Fig 5 Input Data



Consequent	Antecedent	Support %	Confidence %
Capsicum	Corn	75.0	100.0
Corn	Capsicum	75.0	100.0

Fig 6 Frequent Item Sets

VI. CONCLUSION

The DM framework displays the three main characteristics of abstraction, granularity and specificity. The abstraction is evident from the processes of data modeling and data reduction which form part of the framework. The granularity will be proved with the microscopic view of data through the various DM algorithms that will be applied to datasets in order to exploit the specific data characteristics for each of the datasets. The specificity within the DM framework is evident through the formatting of the data for uptake as input to the presentation software. The main aim, however' is for the DM framework to demonstrate its effectiveness in improving the accuracy of crop yield predictions and seed planting recommendations after all the multiple factors are taken into consideration. The DM framework may be given relevance in the future evaluation of other datasets, making it possible for it to be regarded as generic and re-usable. In addition, the new framework as a construct is capable of providing more detailed and granular information to the analyst, thereby enabling conclusions to be drawn effectively.

REFERENCES

[1] Application of Data Mining Framework for the Identification of Agricultural Production Areas in WA Yunous Vagh, Leisa Armstrong, Dean Diepeveen.

[2] G. Fernandez, Data .Mining using SAS applications. Boca Raton: Florida: Chapman & Hall/CRC, 2003.

[3] I.De Falco and A. Della Cioppa, et.al, "An evolutionary approach for automatically extracting intelligible classification rules," Knowledge and Information Systems, vol. (2005), pp. 179-201, 2005.

[4] X. Zhang el al., "CARE: Finding Local Linear Correlations in High Dimensional Data" presented at the 2008 lEEE 24th International Conference on Data Engineering Washington, DC, USA 2008.

[5] J.W.Seifert,"Data Mining: An Overview" in National Security Issues, D. D.Pegarkov, Ed., cd New York: Nova Science Publishers Inc., 2006.

[6] A. Abdullah and I. A. Ansari, "Discovery of cropping regions due to Global Climatic Changes using Data Mining," in CAIR Publications Beijing, 2005,pp.3-11.

[7] A. Abdullah, el al., "Agri Data Mining/Warehousing: Innovative Tools for Analysis of Integrated Agricultural & Meteorological Data," presented at the the IASTED International Conference on Databases and Applications (DBA 2004), Innsbruck, Austria, 2004. ,

[8] J. Han, et al., "Generalization-based data mining in object-oriented databases using an object cube model 11 Data & Knowledge Engineering. Vol. 25, pp. 55-97, 1998.

[9] A. Abdullah, "Analysis of mealybug incidence on the cotton crop using ADSS-OLAP (Online Analytical Processing) tool "Computers and Electronics in Agriculture, vol. 69, pp. 59-72, November 2009.

[10] A. Abdullah, el al., "The case for an Agri Data Warehouse: enabling analytical exploration of integrated agricultural data," 2004.

[11] D. Diepeveen, "Dr.," cd. Perth, 2006, p. Briefing.

[12] J. Boumaa, el al.,"Pedology, Precision Agriculture, and the Changing Paradigm of Agricultural Research" Soil Science Society of America' Journal, pp. 1763-1768, 1999.

[13] D. Gartin and D.Notkin, "Formalizing design spaces: Implicit innovation mechanisms," presented at the Formal Software Development Methods: Int Symposium VDM Europe, 1991.

[14] M. Luck and M. 'D'lnverno, UA Conceptual Framework for Agent Definition and Development, 1I the Computer Journal, vol. 44, pp. 1-20,200l.

[15] M. Dtlnverno, et al., "A formal framework for specifying design methodologies," Software Process improvement Pract" vol. 2, pp. 181-195, 1996.

[16] B. Ullmer and H. Ishii. (2001) Emerging Frameworks for Tangible User Interfaces. Human-Computer interaction in the New A1ileniulIl. 579-601.

[17] D. Gartin, ithe role of formal reusable frameworks, ACU SIGSOFT: Software Eng. Noles, vol. 15, pp. 42-44, 1990.

[18] J. Greenfield and K. Short, "Software Factories: Assembling Applications with Patterns, Models, Frameworks and Tool set in OOPSLA '03, Anaheim, California, USA., 2003

[19] V.Sadowski,"Object-Oriented Application Frameworks," Communications of ACM, vol. 40, pp. 32-38, 1997.

[20] J. Kittler, "Combining Classifiers: A Theoretical Framework," Pattern Analysis & Applic. vol. 1998, pp. 18-27, 1998.

[21] R. Agarwal and M. V. Joshi, I1Pnrule: A new framework for learning classifier models in data mining' University of Minnesota, New York2000.

[22] V. S. Vassilios S. Verykios, et al., "State-of-the-art in Privacy Preserving Data Mining," SIGMOD, vol. 33, pp. 50-57, 2004.

[23] J. Miller and J. Han, Geographic data mining and knowledge discovery 2nd Ed. New York: Taylor & Francis Inc, 2009.

[24] B. Kitchenham, et al., "Experiences of using an evaluation framework" Information and Software Technology, vol. 47, pp. 761-774, 2005.

[25] L. Armstrong. Et al., T2 Data mining can empower growers' crop decision making in T2: Technology and Transformation. 3rd Transforming Information and Learning Conference, Perth, Edith Cowan University, 2007.

[26] Y. Vagh. (2007, The Use Of Data Mining Techniques To Establish Useful Patterns For Crop Variety Selections In A Western Australia Agricultural Context. [Honors].

[27] S.Mitra and T.Acharya, Data Milling: Multimedia, Soft Computing, and Bioinformatics. New Jersey: John Wiley & Sons, 2003.

[28] D. Hand, el al., Principles of Data Mining. Massachusetts: MIT Press, 2001.

[29] H. Cleveland. (1982) Information as Resource. The Futurist. 34-39.

[30] http://agmarknet.nic.in/.

Author's Profile

Dr. J. Rajendra Prasad completed M.Tech., Ph.D and currently working as a Professor & HOD of Information Technology in PVP Siddhartha Institute of Technology. He has more than 30 years of research and teaching Experience. Life member of ISTE and CSI. He published more than 40 Papers in International and National Journal and Conferences. He Author of a Text Book of Mathematical Foundation to Computer Science. He is a referee of one International Journal Published from Rumania. He is guiding five scholars to pursue Ph.D in Computer Science and Engineering. His research areas include Functional Analysis, Cryptography, Data Mining, Networks and Pattern Recognition.

P.Ravi Prakash completed B.Tech.,M.Tech and currently working as a Asst.Professor in Information Technology in PVP Siddhartha Institute of Technology He is having 3 years Teaching experience. His area of interest is computers, Data ware housing and Data Mining. He published one International Journal.

S.Sai Kumar completed B.Tech, M.Tech and currently working as a Asst.Professor in Information Technology in PVP Siddhartha Institute of Technology. He is having 2 year experience. He published two International Journal. His area of interest is Mobile Computing, Image Processing, Middleware Technologies and Computer Networks.

M.SundaraBabu completed B.Tech., M.Tech and currently working as a Asst.Professor in Information Technology in PVP Siddhartha Institute of Technology. He is having 2 year experience. His area of interest is Data Mining and Computer Networks.

K.Swarupa Rani completed B.Tech, M.Tech and currently working as a Asst. Professor in Information Technology in PVP Siddhartha Institute of Technology. She is having 4 ½ years of teaching experience. Her area of Interest is Image Processing and Computer Networks.