IMPLEMENTATION OF COMPUTER INTELLIGENCE ON AGRICULTURE DATASET USING DATA MINING

Pranita Singh[1], Parag Jain[2], Deepak Arya[3], Sakendra Kumar[4], Rachan Karmakar[5]

[1,2,3]Department of CSE, Roorkee Institute of Technology, Roorkee, Uttarakhand, India
[4]Dr. Rajendra Prasad Central Agricultural University, Pusa, Samastipur, Bihar, India
[5]Dept. of Environmental Science, NIMS University, Jaipur, Rajasthan, India
pranitasingh1@gmail.com[1]

Abstract

Fast computerization of scientific techniques and their advancement hasled to huge and continuously increasing amounts of data which makes it very intricate and unmanageable to deduce the facts manually from this data. Therefore, innovative techniques and tools development that support humans in transforming data into valuable knowledge has been the focus of the comparativelynovel and interdisciplinary research area "knowledge discovery in databases".The agriculture sector is the backbone and pillar of India's economy. Despite the emphasis on industrialization, agriculture continues to be a major sector of the Indian economy in terms of its input to gross domestic product (GDP) and a source of employment for millions of people across the country. With the extensive use of remote sensor technology and automated data collection tools, a large amount of agricultural data has been collected and stored in a large database. The comprehension of information meant by such a large amount does a great deal of challenges of relevant research in the field of Agro Informatics and computer intelligence. The current work involves the development of an integrated computer intelligence system to find association rules on agricultural data sets which provides decision making reference.

Keywords: Data mining, Association rule, Land use.

1. INTRODUCTION

Computer Intelligence as we all know is the important area of research in the field of computer science. With the fast technological advancement and broad areas of application, it is becoming universal very rapidly because of its stout applicability in problems, particularly that cannot be solved well by humans as well as in traditional computing structures [1]. Agriculture is such an important sector that it directly employs 30.7 percent of the world's population on 2781 million hectares of agricultural land. Such a strategy does not run smoothly. From sowing to harvest, it faces several challenges. The main issues are the invasion of pests and disease, insufficient chemical application, unsuitable drainage and irrigation, control of weeds, estimation of yield, etc. The application of computers in agriculture was initially reported in 1983 [2]. To solve the existing problems in the agriculture starting from the database [3] to decision support systems, many altered approaches have been recommended [4]. In terms of accuracy and resilience, systems that use computer intelligence are determined to be the best performers out of all of these resolutions. Because agriculture is such a dynamic subject, it is impossible to generalise conditions in order to provide a common solution. These methods have helped us to catch the finer points of each circumstance and offer a solution which is tailored to the challenge at hand. With the advancement in computer intelligence approaches, very complicated issues are gradually being solved.

Data Mining

Data mining assumes a substantial role in quite a huge number of applications [5]. Clementine and DBMiner are some of the extraordinary tools of data mining which are used to perform the task of data mining.

Data mining addresses, as expressed, extraction of stowed away data about forecasting from

enormous records. For the organisations that have been primarily focussing on the main information in the enormous amount of data, this technique is proving to have the extraordinary potential[6]. Tools of data mining anticipate future patterns and ways of behaving, empowering business to go with proactive choices, in view of information. Computerized imminent examination, coming about because of information mining, goes past previous occasion examination by utilizing review devices, trademark for choice emotionally supportive networks. Framework examination is the reason for anticipating key boundaries in characterizing arranging components and their evaluation.

Association Rule Mining

Association rule mining can be simply defined as the process of uncovering interesting and unexpected rules from large data sets. This basically mentions quite a general model that allowsrelationships to be established between different items of the database. An association rule can simply be said an implication or if-then-rule that is supported by data. Agrawal and others were the first to formulate the association rules problem and was named the problem as market-basket problem [7]. Association Rule Mining is one of the major problems in data mining. Study regarding association rule mining has been extended to lot of other areas, such as multimedia data and other areas. And in this study of implementing computer intelligence on agricultural data sets, the technique of association rule mining is used to achieve or results. Association rule mining fundamentallyaims to extract interesting correlations, patterns that occur frequently, associations or casual structures between sets of items in a transactional database or other kinds of data repositories. ARM's main goal is to discover a set of all subsets of elements or attributes that frequently occur in many database records or transactions, and furthermore, extracting rules on how a subset of elementsis affecting the existence of another subset.

The association rule is usually expressed as X→Y,where X is the antecedent and Y is the consequent. Depending on the support and confidence value, the association rule reveals how many times Y has occurred if X has already occurred.

Support: It is defined asprobability with which item or item sets occur in given transactional database.

Support (X) = n(X) / n

where
n:number of transactions in the database
n(X): number of transactions that contains the item set X.

Therefore, support of association rule (X→Y) = support (XUY).

Confidence: It is conditional probability, for an association rule X→Y and defined as

Confidence (X→Y) = support (XUY) / support(X).

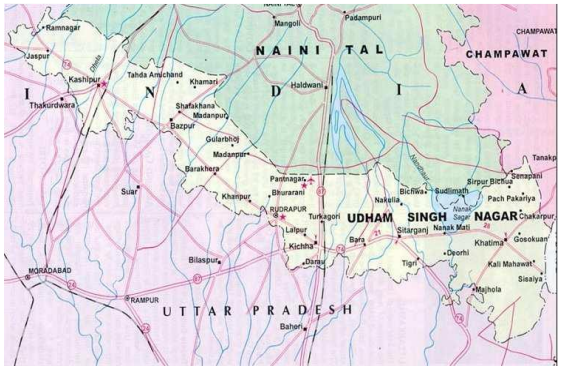
Land use is a result of interactions among society's social foundation, state and its actual necessities from one viewpoint and the regular capability of land on the other [8]. Different examinations have been conducted in various regions Kail Watershed of Central Himalaya [9], Mau region of Uttar Pradesh [10], Doon valley in Dehradun Tehsilof Uttarakhand[11],Kiliyar sub-watershed of Tamilnadu [12] to see the elements of land use utilizing GIS and remote detecting procedures. Notwithstanding work with economical administration of the land, land use data might be utilized for arranging, observing, assessment of advancement, modern action or eclamation. Issues driving area use studies incorporate the expulsion or aggravation of useful land, metropolitan infringement and exhaustion of backwoods. Expanding metropolitan land esteem was a significant driver of farmland improvement,while rising provincial income was an essential driver of transformation of farmland to timberlands and field [13]. Slope gradient was picked as the record of geography to concentrate on their relationship with land use [14].

2. MATERIALS AND METHODS

Study Area

This study was conducted in the Udham Singh Nagar area, located in Uttarakhand state of India. TheUdham Singh Nagar district is located in the Tarai district of Kumaun division. Nainital and Champawat districtsborders the region from north. Moradabad, Rampur, Bareilly and Pilibhit districts of Uttar Pradesh borders the area fromsouth, Bijnor district of Uttar Pradesh

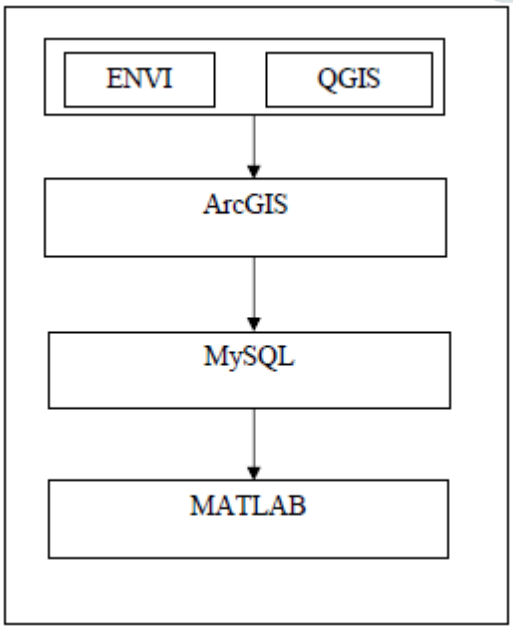
from west, Nepal from the east. The international boundary between India and Nepal is formed by Sarada River.



Location of study area

Software used

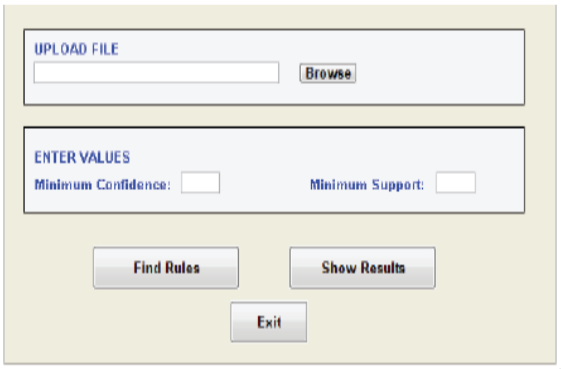
Below are the software which are used at different levels of implementation. Pre-processing and changing data to shape file format is done by ENVI software and also to change the. QGIS is used to derive slope from DEM and to obtain the shapeSfile of slope.ArcGIS is used for analysis of data and merging of thetwo databases to have a single .dbf file containing attributes of both the databases. After that MySQL is used for converting the database to transaction database by adding attributesand updating the database by applying SQL queries. Association rule mining is thenapplied on this transaction database for finding association rules. Association rule miningprogram based on Apriori is coded in MATLAB.



Software used

3. RESULTS AND DISCUSSION

Below shown is the developed GUI in MATLAB which is used for mining agricultural datasets.



Developed GUI

Most of the types of land uses are concentrated at the slope level zero. In the study area that we have taken, we have eight kinds of land use types.The supports of the various land use types at this slope level varies between 0% and 52%. The support for fallow land is found to be 51.92% which is the highest support value of all that implies a large amount of area is left as it is for a period to retain its fertility. The land use dataset which we have used is of a particular season. Therefore, the fallow is showing a high support because after one crop land is left idle for a period of time. This land can be utilized in an efficient way to have a high production. The results for agricultural land are slightly similar to the results of the case study of Mau District, Uttar Pradesh. In Mau district, agriculture is the major land use kinds due to the one of fertile soil of the world.

It is clear from this study that agricultural production in this area is mainly concentrated on the land having lower slopes. At lower slope levels, built up land use type has a very high confidence which clearly shows that the population of the area under research is mainly resides in the low slope areas. Means residential area is basically the low slope area.

4. CONCLUSION

The objective of the study was implementing computer intelligenceto mineagricultural datasetsusing MATLAB. ENVI, ArcGIS and MySQL softwares are used for pre-processing and transformation of agricultural datasets of land use and slopeobtained from Agro meteorology department ofAgriculture College, G.B.P.U.A. & T., Pantnagar. GUI that incorporated Apriori algorithm developed in MATLAB is successfully

generating association rules out of land use and slope databases. It was found thatall land use types mainly concentrate on the landhaving lower slopes. Huge portion of land has been found fallow and this land type can be used efficiently in having good agriculture outputs in terms of production. Therefore, productionin this area is mainly concentrated on the landhaving low slope values. The forest area which is also an integral part of agriculture is most suited with the higher slopes, thereby increasing the forest area which plays a vital role in binging of the soil.

Theresults obtained are expected to provide help and a good assistancefor land use planning. In future, similar work canbe performed for providing decision makingreference regarding land use planning in differentcatchment areas of different part of the country. The MATLAB has been proved to be a strong tool for implementation of algorithm and in the development of GUI. Other algorithms can also be applied to see the variations in the outcome and can be used for a good comparison. Other than MATLAB, algorithm can also be implemented with other programming languages to see the performance variation.

REFERENCES

[1] E. Rich and Kevin Knight. "*Artificial intelligence*", New Delhi: McGraw-Hill, 1991.

[2] D.N. Baker, J.R. Lambert, J.M. McKinion, ―GOSSYM: A simulator of cotton crop growth and yield,‖ Technical bulletin, Agricultural Experiment Station, South Carolina, USA, 1983.

[3] P. Martiniello, "Development of a database computer management system for retrieval on varietal field evaluation and plant breeding information in agriculture," *Computers and electronics in agriculture*, vol. 2 no. 3, pp. 183-192, 1988.

[4] K. W. Thorpe, R. L. Ridgway, R. E. Webb, "A computerized data management and decision support system for gypsy moth management in suburban parks," *Computers and electronics in agriculture*, vol. 6 no. 4, pp. 333-345, 1992.

[5]Han, J. and Kamber, M. (2006). *Data Mining: Concepts and Techniques*.2nd ed. The Morgan Kaufmann Series in Data ManagementSystems, Jim Gray, Series Editor Morgan Kaufmann Publishers.

[6] Milovic, B. and Radojevic, V.Application of data mining in Agriculture. *Bulgarian Journalof Agricultural Science, **21 (1)** 2015, 26-34. Agricultural Academy*

[7] Agrawal, R. and Srikant, R. (1994). Fast algorithms for miningassociation rules. In: Proceedings of 1994 InternationalConference VLDB, Santiago, Chile. pp. 487-499.

[8]Balak, R. and Kolarkar, A.S. (1993). Remote sensing application inmonitoring land use changes in arid Rajasthan. *Inter. J. RemoteSens.* **14(17)**, 3191-3200.

[9] Rawal, H.S., Rawat, J.S., Kumar, M., Pant, N.C. and Rani, N. (2014).Land use/cover dynamics of Kail Watershed, Central Himalaya,India using Remote Sensing and GIS techniques. *Inter. JournalAdv. Remote Sens. GIS Geog.* **2(2)**, 55-59.

[10] Singh, P. and Singh, S. (2011). Landuse pattern analysis using remotesensing: a case study of Mau District, India. Scholars researchlibrary. Archives of Applied Science Research, **(5)**, 10-16.

[11] Tiwari, K. and Khanduri, K. (2011). Land use / land cover changedetection in Doon valley (Dehradun Tehsil), Uttarakhand: usingGIS & remote sensing technique. *Inter. J. GeomaticsGeosci.*, **2(1)**, 34-41.

[12] Poongothai, S., Sridhar, N. and Shourie, R.A. (2014). Change detectionof land use/ land cover of a watershed using remote sensing andGIS. *Inter. J. Engg. Adv., Tech.*, **3(6)**, 226-230.

[13] Li, M., Wu, J. and Deng, X. (2013). Identifying drivers of land usechange in China: A spatial multinomial logit model analysis.*Land Eco.*,**89(4)**, 632-854.

[14] P.Surya. (2018). Crop yield prediction in agriculture using data mining predictive analytic techniques. *International Journal of Research and Analytical Reviews*. **5(4)**, 783-787.