

Application of data mining in Rural Planning

Abolfazl Shahbazi, Maryam Karambeygi

M. Sc. graduate, Rural Planning, Islamic Azad University, Iran

M. Sc. graduate, IT Engineering, Islamic Azad University, Iran

Abstract— Information technology is one the world's most important development aspects and many countries consider the development of information technology as one of the most vital aspects of infrastructure development. This technology can offer vast opportunities and capabilities to rural areas which will in turn aid in effectively solving many problems associated with such areas. Information in the agricultural field and rural development is not only crucial but the maintenance and analysis of data is considered one of the main subjects in efficiency and capital improvement. Strengthening data collection and analysis provides perhaps the greatest opportunity for researchers and policy makers for rural areas. Utilization of such data is considered a driving force that improves the economic dynamism while creating a new type of knowledge-based economy. The main purpose of this paper is to assess the application of data mining in rural planning and management with the goal of achieving sustainable agricultural and rural development while promoting productivity in the agricultural field.

Index Terms—Data Mining, Business intelligence, Rural Planning

I. INTRODUCTION

With regard to government duties in the third millennium and heavy competition in the area of economics throughout the world, the countries that are able to utilize modern technologies in their production areas which includes rural areas will be successful in managing their society effectively. Today, information technology and communications is considered the core of rural management in countries such as Canada, the United States and South Korea. Nowadays, the development of societies and institutions are due to the advancements of appropriate infrastructures namely information technology and communications such that the management of these infrastructures is independent from geographical locations and distances from centers of civilization, capital, and equipment and only by relying on knowledge and management capabilities. The benefits of rural ICT development have various dimensions. In the field of rural economics, ICT can entail evolution in the agricultural, livestock, and handicrafts industries. By using modern scientific methods, the background is created for knowledge based activities and systematic thinking. Strategic management activists in the rural sector can obtain information on internal and foreign markets by accessing organized databanks therefore they are able to make suitable decisions concerning the development of such areas using correct analysis. Strategic management for the rural sector in developed countries is one of the main issues concerning the

modern science of data mining such that today, this science has many applications for areas such as citizen segmentation, transportation and traffic planning, waste management, education planning, increasing citizen satisfaction, locating urban facilities, citizens relations management, citizens requirements identification, staff evaluation, rural planning, human resources management, and various other areas. Since data mining is a process that utilizes various modeling and analysis techniques to identify data patterns and relationships for adequate predictions, managers of the rural sector are therefore aided in various aspects of their jobs. Utilizing the modern science of data mining as the prevalent method for problem solving has been crucial for rural area management's data bases such that today, data mining plays a vital role in all micro and macro planning issues.

II. DATA MINING

The main reason for data mining becoming the center of attention in the information industry is its accessibility in obtaining vast volumes of data and the need to extract beneficial information and knowledge from such data. The resulting information and knowledge is applicable in business management, productivity control, market analysis, engineering design, and scientific research. Data mining may be considered as the result of information technology evolution. This evolution stems from the evolution in the data industry namely data gathering, database creation, data management, and data analysis. The expression "data mining" is used by statisticians, database researchers, management information systems and business communities. Knowledge identification in databases generally refers to the overall process of identifying beneficial knowledge from databases. Activities such as data preparation, data selection, data cleansing and correct understanding of the data mining process enables the extraction of useful information. Data mining originates from traditional data analysis and statistical approaches and consists of analytical techniques from other scientific branches. Data mining is described in the following five basic operations.

A. Predictive modeling and classification

This data mining operation is used to predict a specific event. This operation assumes that the analyzer has a certain question in mind. In response, the model returns a rating of the probability of certain events taking place. For example, if a rural analyzer aims to predict the number of farmers leaving the area, they should enter two types of data into the data mining tool.

- Data relating to the number of residents that have previously left the area. This data is known as "bad" data.
- Data relating to number of residents that have stayed in the area and are long term residents. This data is known as "good" data.

The data mining tool checks these data to define the variables relevant to residents that have left the area and residents that currently reside in the area. For example, the results of the analysis may be presented in such form: "a male village resident, who is the household head of six members, is over 40 years old, has an annual average income of over 15 million rial, and is a home owner has a probability of 35% of leaving." Examples of investigation questions for predictive data mining are related to issues that include dependencies, patterns, working procedures, tendencies, and principles. For example:

- Which recommendations motivate residents to expand their agricultural activities? (tendency/working procedure)
- Which village residents should be targeted for certain products? (dependency)
- What are the signs of unethical activities in expanding village agriculture and livestock? (pattern)
- Which farmers are faced with fewer risks in terms of losses and negative returns? (principle)

B. Links analysis

The links analysis operation in data mining consists of a set of mathematical algorithms and visual techniques that identify links between records within a database. Results are presented as images. This analysis is used to identify dependencies and sequential patterns in data mining. For example, links analysis can determine which agricultural products are normally bought together (e.g. grapes and apples). Another example is which products are normally placed alongside one another in the cart of goods so based on the results, farmers can produce supplementary products.

C. Database classification

This data mining operation consists of a set of algorithms that classifies similar data. This is a clustering technique for data mining. This classification is usually the first step in data selection and takes place before other data mining operations. For example, sectioning may divide village residents into two groups consisting of people that use information technology on a regular basis and those that use such technologies coincidentally. Another example is classifying village residents into two groups consisting of residents that are in favor of reconstructing village houses and those who are in favor of maintaining traditional rural architecture and texture.

D. Deviations detection

Detecting deviations in data mining consist of a set of algorithms that determine records outside of what was predicted and present reasons for such abnormalities. Deviations detection normally takes place to identify illegal

activities. The illegal acquirement of land for example where data related to farmers information, amount of products produced, irrigation levels etc. in specific periods of time is used to predict the probability of acquiring land illegally. Abuse of natural resources and loss of groundwater for the cultivation of rural products is another example where such operation is used for predictive purposes. The use of hazardous chemicals for the environment to increase agricultural and livestock production is another example. Results obtained from this operation may be presented as follows: Farmers with an income of over 8 million rial in the previous month who have purchased water resources and obtained agricultural loans of over 60 million rial have a probability of 65% of their productions being of less quality due to their use of fertilizers.

E. Dependency detection

This data mining technique is used to determine behavior processes or specific events. Dependency detection is used to link events. For example, results may be presented as "the probability of customers to purchase expensive grapes and raisins in the amount of over 25 ton, is three times higher than customers that purchase these products with less quality". Dependency detection is based on the following concepts: An analyzer of the rural sector aims to assess specific customer purchasing trends in order to improve returns on investments in the agricultural field (e.g. wholesalers of two urban areas) and design marketing plans. First, the analyzer uses defined characteristics of customers and attempts to classify them. Then, purchasing behaviors are studied. The analyzer continues this process until the appropriate final classification is defined. Data mining capabilities are not something that is easily available and affordable. For data mining it is necessary to create a decision-aiding application or specifically a data mining application using a data mining tool. This data mining application may utilize a combination of classic or advanced elements such as artificial intelligence, pattern determination, databases, traditional statistics and graphics to present latent relationships and patterns in gathered data. Data mining is used to analyze data to discover beneficial latent information.

II. DATA MINING APPLICATIONS

It should always be considered that with the existence of all technologies, rural industries should have a fundamental dependency on data mining to justify the related costs and time consumption. Of the most common driving forces for data mining utilization is to increase rural products market shares. Methods for achieving such goal include the introduction of new products and attaining competitors market shares. In both cases, data mining application can aid in decision making to reach ones goals. Various types of data mining applications exist. Five of the most common applications are introduced below.

A. Market management

Marketing management is the organizational discipline which focuses on the practical application of marketing orientation, techniques and methods inside enterprises and organizations and on the management of a firm's marketing resources and activities.

Examples of the most common data mining applications are related to market management are introduced below:

- Cross-selling: cross-selling opportunities for customers and sections of the market where the probability of purchasing local products and services due to advertisements is defined.
- Advertisements and campaigns: natural market classification is recognized. For example, key sale time intervals for agricultural, gardening, handicrafts and livestock products are identified and campaigns are presented accordingly.
- Goods cart analysis: to identify the village products people buy together from shops. This information is used to classify products and their time of layout in shops or to advertise and price items.

B. Immigration management

- Immigrating villagers: by using vulnerability analysis, it is determined which villagers may immigrate. Therefore, strategies may be designed to prevent residents from immigrating.

C. Fraud detection

- Fraud and bribery: model extraction and clarification of circumstances for granted quota to the rural section which seem suspicious or are indications of fraud such as sales of state agricultural equipment and machinery in the free market. This model is used to predict the reliability of new applicants.
- Insurance fraud: large volumes of data concerning rural claims are analyzed to detect fraud related to health insurance, crop insurance and rural insurance.

D. Rural products distribution

- Inventory control: predictive models may be designed to predict the time and place that rural products are needed. By using such models, inventory control and distribution of rural products are improved.

III. DATA MINING ACTIVITIES STEPS

Data mining steps do not need to take place in a certain order. Figure 1 shows activities that can take place simultaneously. The following list describes data mining activities concisely.

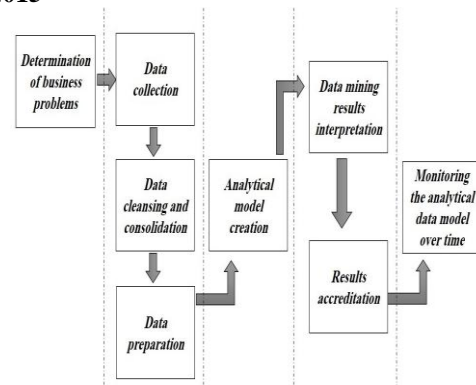


Fig 1- Data mining activities steps

1) Determination of business problems

Before carrying out data mining, the objectives should be defined and prioritized (e.g. increasing profits, decreasing costs, designing modern strategies, presenting rural products, increasing local production market shares and decreasing imports). Financial and time investments are needed to achieve each of these goals. The details of this step are presented in the figure 2

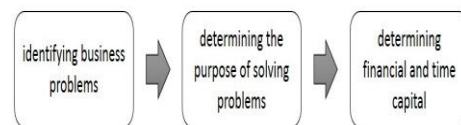


Fig 2- Determination of business problems activities steps

2) Data collection

One of the most time consuming tasks in data mining is collecting adequate data. In order to do this effectively, all relevant data for analysis should be gathered. This data includes stored data in government organizational operating databases such as agricultural machinery development agencies, rural cooperative companies, rural services companies, agricultural water management, registration information, villagers personal information, rural architecture, rural guidance plans and any other relevant data. When the data sources are identified, all relevant data elements are extracted. The details of this step are presented in the figure 3.

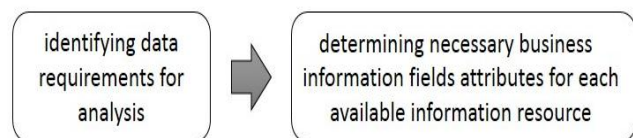


Fig 3- Data collection activities steps

3) Data cleansing and consolidation

Data from various sources should be cleansed and consolidated. If it is necessary to complete external data with internal data, the external data should match internal data and appropriate contents should be defined. Cooperation with software and IT engineers is effective at this stage. The details

of this step are presented in the figure 4

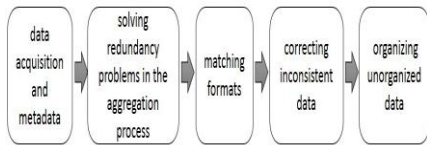


Fig 4- Data cleansing and consolidation activities steps

4) Data preparation

Before creating an analytical model, data should be prepared. A part of data preparation is classifying variables. Variables may be discrete or continuous, qualitative or quantitative. Uncertain variables must be eliminated or replaced with probable values. Knowing the minimum, maximum, mean, median, and mode of quantitative variables provides adequate insight for data. To simplify the preparation process, data reduction transforms should be applied. The aim of data reduction is to combine a few variables into one variable for controllability analysis. For example, education levels, income, marital status and postcodes are combined. The details of this step are presented in the figure 5

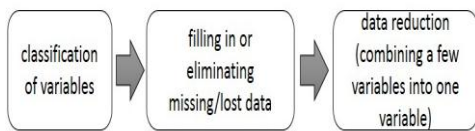


Fig 5- Data preparation activities steps

5) Analytical model creation

One of the most prominent activities in data mining is to create an analytical model. An analytical data model shows an integrated, comprehensive, and time-dependent data structure consisting of internal and external data sources that are preprocessed. Upon implementation, this model should be able to continue the "learning" process such that data mining tools are utilized regularly and assessed by data mining experts. In other words, results of each analysis should be considered as new inputs and used in later analysis. The details of this step are presented in the figure 6



Fig 6- Analytical model creation activities steps

6) Data mining results interpretation

When data mining is implemented and results are obtained, the main responsibility is interpreting these results. Two points must be considered at this stage: how easily can the results be influenced and can these results be persuasive enough to present to the ministry of agriculture offices. The details of this step are presented in the figure 7

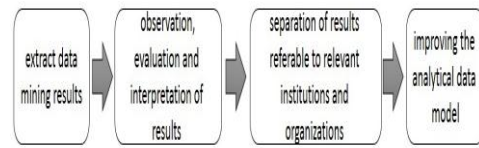


Fig 7- Data mining results interpretation activities steps

7) Results accreditation

Results should be compared with other published industrial statistics. Deviations from these statistics should be determined along with the reasons for such deviations. Up to date industrial statistics should be used since these statistics vary over time. Data collection criteria should be compared with data collection criteria relevant to these statistics. Time periods for resulting data and industrial statistics time periods should be compared. The created model data selection criteria with relevant time period should resemble industrial statistics data selection criteria and time periods. The details of this step are presented in the figure 8



Fig 8- Results accreditation activities steps

8) Monitoring the analytical data model over time

Industry statistics are usually obtained using very larger samples. Therefore it is important to accredit the analytical data model with these statistics at regular time intervals. Industrial statistics change over time and some industries have seasonal variations. The details of this step are presented in the figure 9.

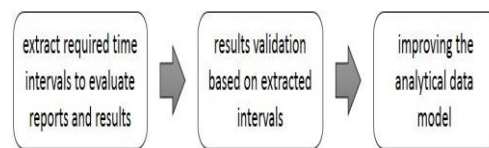


Fig 9- Monitoring the analytical data model over time activities steps

IV. CONCLUSION

As previously stated, strengthening data collection and analysis is a fundamental necessity for rural areas policymakers such that efficient utilization of the resulting opportunities will entail society's economic dynamism while creating a knowledge-based economy. In other words, the data mining process of identifying appropriate data, strengthening data collection, building a data warehouse, data analysis and subsequent decision making will have tremendous impact in the progression of societies especially in third world countries. Data collection and adequate data analysis has many applications in various areas including market management, agricultural products, immigration management, rural products distribution methods, fraud

detection concerning rural areas facilities etc. Finally, this paper describes data mining steps.

REFERENCES

- [1] Corden, Max, and J. Peter Neary. 1983. "Booming Sector and De-Industrialization in a Small Open Economy." *Economic Journal* (London) 92:825–48.
- [2] Field, Donald R., and William R. Burch Jr. "Rural Sociology and the Environment. 2nd ed. Middleton", WI: Social Ecology Press, 1991.
- [3] Krannich, Richard S., and A. E. Luloff. "Problems of Resource Dependency in U.S. Rural Communities." *Progress in Rural Policy and Planning* 1:5–18, 1991.
- [4] Rob Gerritsen, Assessing Loan Risks: A Data Mining Case Study, *IEEE IT Pro* November | December 1999.
- [5] Mukalula Poverty Alleviation through construction industry: The case of Zambia's rural areas. Second LACCEI Conference, Miami, Florida, 2004.
- [6] Bender, L. D., B. L. Green, T. F. Hady, J. A. Kuehn, M. K. Nelson, L. B. Perkinson, and P. J. Ross. "The Diverse Social and Economic Structure of Non-Metropolitan America. Washington, DC: Agriculture and Rural Economic Division, Economic Research Service", U.S. Department of Agriculture, 1985.

AUTHOR'S PROFILE

Abolfazl Shahbazi was born in 1987 and has graduated with a B.Sc. degree in Civil Engineering from Arak University, Iran in 2009. He then graduated with a M.Sc. degree in Rural Planning at Islamic Azad University, Yadegar Emam branch, Tehran, Iran in 2014. His M.Sc. project was about the rural strategies on developed countries. He works in Economic Development and Culture Division. He has over 4 years' experience in Rural and City planning.

Maryam Karambeygi was born in 1988 and has graduated with a B.Sc. degree in Computer Engineering from Arak University, Iran in 2012. She then received a M.Sc. degree in IT Engineering at Islamic Azad University, Ghazvin branch, Ghazvin, Iran in 2015. Her His M.Sc. thesis is titled "Active view prediction, to store aggregated data in data warehouse" which concerns data warehouse and its uses. She currently works in BI and Data Mining projects.