Bulgarian Journal of Agricultural Science, 21 (No 1) 2015, 26-34
Agricultural Academy

APPLICATION OF DATA MINING IN AGRICULTURE

B. MILOVIC[1] and V. RADOJEVIC[2]
[1] *Agricultural Enterprise "Sava Kovačevic" at Vrbas, 21460 Vrbas, Serbia*
[2] *University of Novi Sad, Faculty of Agriculture, 21000 Novi Sad, Serbia*

Abstract

MILOVIC, B. and V. RADOJEVIC, 2015. Application of data mining in agriculture. *Bulg. J. Agric. Sci.,* 21: 26-34

Today, agricultural organizations work with large amounts of data. Processing and retrieval of significant data in this abundance of agricultural information is necessary. Utilization of information and communications technology enables automation of extracting significant data in an effort to obtain knowledge and trends, which enables the elimination of manual tasks and easier data extraction directly from electronic sources, transfer to secure electronic system of documentation which will enable production cost reduction, higher yield and higher market price. Data mining in addition to information about crops enables agricultural enterprises to predict trends about customer's conditions or their behavior, which is achieved by analyzing data from different perspectives and finding connections and relationships in seemingly unrelated data. Raw data of agricultural enterprises are very ample and diverse. It is necessary to collect and store them in an organized form, and their integration enables the creation of agricultural information system. Data mining in agriculture provides many opportunities for exploring hidden patterns in these collections of data. These patterns can be used to determine the condition of customers in agricultural organizations.

Key words: Data mining, agriculture, data processing, information systems, agricultural enterprises

Introduction

In today's conditions agricultural enterprises are capable of generating and collect large amounts of data. Growth in data size requires automated method to extract necessary data. By applying data mining technique it is possible to extract useful knowledge and trends. Knowledge gained in this manner, may be applied to increase work efficiency and improve decision making quality. All of this indicates there is a need for a new generation of computer theories and tools to help people in extraction of useful data from constantly growing pool of digital data. Information technology has become an integral part of our daily life. Techniques for managing data have become necessary and common in industry and services. Improvements in efficiency can be achieved in almost every aspect of business. This is especially true for agriculture, in order to modernize and better apply GPS technology. Agricultural companies in addition to reaping the fruits on the fields have started collecting large amounts of data. Large amount of information about soil and crop properties, which enables higher operational efficiency, is often contained in these data – in order to find this information it is necessary to apply adequate techniques. This is a common problem from which arises the term data mining. Data mining techniques are directed towards finding those schemes of work in data which are valuable and interesting for crop management (Mucherino and Ruß, 2011). Since the files in question are large (Miller et al., 2009), information is often hidden, so data mining techniques are used for their detection. Data mining uses different techniques for dividing the data in subsets, predicting revenue based on data, grouping data in subgroups or defining results for some dimensions of data. Agriculture has developed from mule and plow into high-tech business. Data mining techniques have made statistics enhanced by quality of data from collection to evaluation. Data mining reveals hidden information which agricultural management uses for improving its decisions. By ease of use and the possibility of presenting complex results in a simple fashion, data mining has shown to be fertile ground for future inovation in the field of agricultural statistics (Miller et al., 2009).

E-mail: boris.milovic@gmail.com; radojevicvuk@yahoo.com

Definition of the Term Data Mining and Knowledge Discovery

Simple definition of data mining in marketing is (Foss and Stone, 2001): extraction of previously unknown, understandable and adequate information from large data storages and their use for key business decisions in order to support them are carried out, formulating tactical and strategic marketing initiatives and measuring their success. This definition emphasizes the following aspects of 'Data Mining': information extraction, large repository of data and formulating initiatives. 'Data mining' (Buttle, 2009) simplifies finding correlation in the data. 'Data Mining' is extremely interdisciplinary field. It covers fields of databases, expert systems, information theories, statistics, mathematics, logics and a series of connected fields. Fields in which data mining can be successfully applied are diverse, for example organization business, economics, mechanics, medicine, genetics. Generally speaking, data mining is applicable in all fields in which certain regularities, connections and rules are to be found from large data.

It can be said that 'Data Mining' is finding rules in data. Technology of data mining is narrowly connected to data storage and is intertwined with database management system. Data mining involves the process of finding large quantity of previously unknown data, and then their use in important business decision making. Key phrase here is 'unknown datum' which means that the datum is buried in large quantity of operational data which, if analyzed, provide relevant information to organizational decision makers.

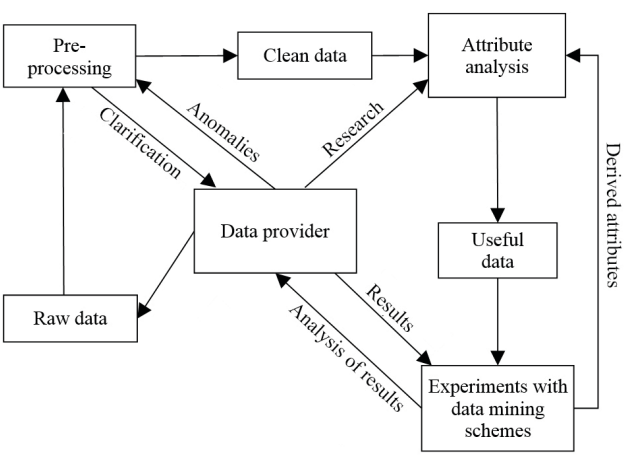

Fig. 1. Process model for a machine learning (data flow diagram)

In the model of Figure 1 data flow is circulating clockwise. Raw data are shown as a single table that is required by the data mining algorithms included in WEKA – an open source system that provides a collection of visualization tools and algorithms for data analysis and predictive modeling. Table is then transformed into attribute/value table that includes header information based on the data types of attributes. This requires cleaning of data, dealing with missing values, discovering false values, etc. After this data is processed by the data mining algorithms, it is transformed into form that will produce readable, accurate data model. One or more versions of clean data are now processed by the data mining schemes. It is now decided which groups of output data are new and accurate enough or interesting to justify further research, and which are common knowledge for that field.

ICT technologies are more and more implemented in agricultural enterprises in order to respond to the needs of agronomists and managers in their daily decision making activities. Data mining tools can be very usefull to control human limitations such as subjectivity or fatigue induced errors, and to provide indications for decision making processes. The essence of data mining is in the identification of relationships, patterns and models which provide support to predictions and dicision making process for agrotechnical measures and decisions in management or sales. Those models can be called predictive. These models can be integrated in the information systems of enterprises as decision making models, reducing subjectivity as well as decision making time. Furthermore, the use of information technologies in agriculture enables comprehensive management of agricultural knowledge and its safe exchange between the recipient and provider of agricultural services. Wide use of information technologies (IT) enables the elimination of manual tasks of extracting data from charts or filling in of the specialized questionnaire, extracting data directly from electronic records, transfer to secure electronic system of agriculturual documentation which will reduce the cost of agricultural products. Return of information through computers can help the decision making quality and avoiding human error. When there is a large amount of data to be classified, human decision making is mostly poor. Data mining represents the process of finding raw data with the help of computers and extraction of their meaning. It is often defined as finding previously unknown potentially useful information from large amount of (unstructured) data (Milovic, 2011). Thanks to this technique, it is possible to predict trends or user behaviour and in that way secure business success of the enterprise. It is achieved by analysing data from different perspectives and

finding connections and relationships in seemingly unrelated information. Once discovered knowledge when presented to the user, evaluation measures may be improved, mining can be further 'refined', new data may be selected or further transformed, or new data sources may be integrated, with the aim of gaining different, corresponding results (Zaïane, 1999). In the process of data mining previously unknown trends and patterns are discovered from the base of historical information and those information turns into significant business solutions (boirefillergroup.com, 2010).

When tasked with knowledge integrity evaluation, two biggest challenges are (Yang and Wu, 2006):

How to develop efficient algorithms for comparing the contents of two versions of knowledge (before and after evaluation). This challenge requires the development of efficient algorithms and data structures for knowledge integrity evaluation in a dataset.

How to develop algorithms for evaluation of influences of certain modifications to the data on the statistical development of individual patterns acquired by general classes of data mining algorithms. Algorithms are developed here which measure the influence of modifications of data values on discovered statistical meaning of patterns, although it would be unfeasible to develop a universal measure for all data mining algorithms.

Knowledge gained in this manner can be applied in corresponding area to increase work efficiency and improve the decision making quality. These points to the fact that there is a great need for new generation of computer theories and tools to help people in extraction of useful information from constantly growing digital data (Fayyad et al., 1996). Type of analysis guided by data is used because analysis guided by interest can overlook unexpected patterns in data. Rules of association are mostly used in this type of analysis. This approach which uses both types of analysis has its upsides and downsides, because users are not excited about vast number of findings far beyond their field of interest, yet unexpected patterns don't stay unseen. Data mining techniques are used for finding data, classification and extraction of information from large files. These techniques, often used in private sector for market research, detecting fraud and customer relations management (CRM), can also be used by the statistics agencies, for analyzing their files. Since large files are characteristic for many statistics agencies, data mining techniques were not used to a large degree for improving official statistics (Miller et al., 2009).

Data Mining Methods

Data mining represents, as stated, extraction of hidden information about predicting from large files. This is a new technology with great potential to assist companies focusing on the most important information in their large data. Tools for data mining predict future trends and behaviours, enabling business to make proactive decisions, based on knowledge. Automated prospective analysis, resulting from data mining, goes beyond past event analysis by using retrospective tools, characteristic for decision support systems (Veenadhar et al., 2011). System analysis is the basis for predicting key parameters in defining planning elements and their quantification (Novkovic et al., 2011).

Data mining uses different methods for the purpose of acquiring necessary information. Different methods are used for different purposes, where every method has its advantages and drawbacks. Data mining tasks can be divided into descriptive and predictive. While descriptive tasks aim to find human interpreted patterns and associations, after considering data as well as the entire model construction, predicting tasks aim to predict some response of interest. Even though the goals of description and prediction may overlap, the main difference is that predictive tasks require a special variable of response. Response may be categorical or numerical, which further classifies data mining as classification and regression. Predictive tasks enable prediction of variable value based on other existing information. Descriptive tasks, on the other hand, unify data in a certain manner. Main predictive and descriptive tasks of data mining can be divided into (Weiss and Davison, 2010):

Classification and regression – identification of new patterns with previously defined goals; These tasks are predictive and involve the construction of models in the aim of predicting a goal, or dependent, variable from the set of explained, or independent variables. Classification is the process of finding the function for classifying data in one of several classes. For tasks of classification the goal variable usually has a small number of discrete values, while with tasks of regression the goal variable is continuous.

Association rules – analysis of rules of association is a descriptive data mining task which involves determining patterns, or associations, between the elements in the dataset. Associations are presented in the form of rules, or implications.

Cluster analysis – descriptive data mining task in which the goal is to group similar objects in the same cluster and different in different clusters. Process of grouping defines

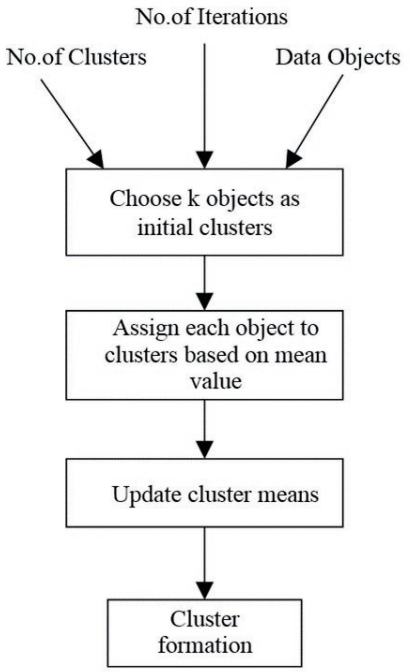

Fig. 2. Cluster formation using k-means (Rajesh, 2011)

groups of similar data, but different than the rest of data. In this process variables by which the grouping is most efficient are often identified. K-means algorithm selects (k) number of objects randomly, where each object represents a cluster mean. Each object is assigned, based on the distance between the object and the cluster mean, to the cluster to which it is most similar. Then it computes new mean for each cluster. Algorithm repeats itself until it reaches its goal. Example is represented in Figure 2.

Text mining tasks – most of available data is in the shape of unstructured or partially structured text, different from conventional data, which is completely structured. Text is unstructured if there is no previously defined format, or structure in the data. Text is partially structured if there is a structure connects to one part of data. While text mining tasks usually fall in the category of classification, clustering and association rule mining, it is best to regard them as separate, because unstructured text requires special consideration. Method for presenting textual data is especially critical.

Link analysis – form of network analysis examining associations between objects. Link classification predicts the category of an object, not just based on its characteristics, but also on the basis of existing connections, and the characteristics of objects connected by a certain path (Getoor, 2003). With the completion of information analysis, all results are presented in a clear manner, often in the form of charts or diagrams either two-dimensional or three-dimensional. Software even enables the user to change variables and view the effects in real time.

Data Mining Techniques

Analytic techniques used in data mining are in most cases well known mathematical techniques and algorithms. Although data mining is a young technology, the process of data analysis alone doesn't include anything new. The fact that connected those techniques and large databases was the cheapening of storage space and processing power. Data mining techniques are used to find patterns, classify records, and extract information from large datasets. In large datasets information is usually hidden, but data mining techniques can be used to discover, and transform in useful knowledge. By the use of data mining tools in a spreadsheet of data analysis software for identifying patterns and connections, it is possible to profile customers and develop business strategies (Milovic, Complexity of implementing the concept of CRM, 2012). Data mining techniques are mainly divided in two groups, classification and clustering techniques (Ramesh et al., 2013). Different techniques are used in practice to classify data in subsets, predict outcomes based on data, group records in subgroups, or score tendencies for certain measures in the records (Miller et al., 2009). There are several applications of Data mining techniques in the field of agriculture. Some of the data mining techniques are related to weather conditions and forecasts. For example, the K-Means algorithm is used to perform forecast of the pollution in the atmosphere, the K Nearest Neighbor (KNN) is applied for simulating daily precipitations and other weather variables, and different possible changes of the weather scenarios are analyzed using SVMs (Ramesh and Vishnu, 2013). Some of the data mining techniques are:

Artificial neural networks are analytic techniques formed on the basis of assumed learning process in the human brain. The same way the human brain after learning process is capable of deducing assumptions based on earlier observations, neural networks after learning process are capable to predict change and events in the system. Neural networks are a group of connected input/output units in which every connection has its weight. Learning process is conducted by balancing the network based on connections existing between elements in the

examples. Based on importance of cause and effect between certain data stronger or weaker connections are formed between 'neurons'. Network formed in this manner is ready for work on unknown data and reacts based on previous knowledge. Conducted research (Štasnyet al., 2011) tested the influence of multilayer neural networks on crop yield prediction, as well as comparing the precision of this approach with well-known regression model designed for predicting empirical data. The point is to apply the neural network technique on tasks which are not solved easily by using nonlinear regression, prediction and classification methods. Application of multilayer neural network has shown to be very precise in predicting crop yield success compared to regression model.

Condition tree is a graphic representation of relations which exist in the database. It is used for data classification. Result is displayed in the form of a tree, hence the name of this technique. Decision trees are mostly used in classification and prediction. It is a simple yet powerful way of presenting knowledge. Models constructed from condition tree are presented in the form of tree structure. Cases are classified by sorting them from root node to the leaf node. Nodes branch are based on if-then conditions. Representation in the form of a tree is clear and easy to understand, and condition tree algorithms are significantly faster than neural networks and their learning is much shorter. Condition tree model uses an algorithm to separate the data based on maximal difference found in a set of variables in regard to desired variable. Division is carried out in the form of sequences. Rule of division finds a variable and the value in which it is divided, and proceeds downward along the corresponding branch to the node. On every new nodem, rule of division will repeat the procedure, observation inside that node is sent downward to the next node. The end nodes are called leaves (Miller et al., 2009). Condition tree is a tree where every node which is not a leaf represents a test or decision about data item currently considered. The choice of a certain branch depends on the outcome of the test. To classify specified datum, we start from the root node and follow the claim downward to the end node, by which point we make a decision. Decision tree can also be interpreted as a special form of a set of rules, characterized by its hierarchical organization of rules. Condition tree technique can be applied in solving problems of crop management, agricultural security, animal husbandry, food safety, water, waste management etc (Huang et al., 2010).

Genetic algorithm is a method of optimization and research technique which uses techniques based on evolution, specifically inheritance, mutation and selection. Genetic algorithm at the same time processes a set (population) of potential solutions (individual) for a given problem. Algorithm begins with a set of solutions called sub-population; more specifically it creates a certain number of random solutions. All those solutions don't need to be good, a set of solutions may be completely omitted, or the solutions may even overlap. Suitability of those solutions based on criteria of performance is evaluated and used to select solutions making a new, better subset of potential solutions (Huang et al., 2010). Bad solutions are disregarded, and good ones kept. Good solutions are then hybridized and the whole process is repeated. In the end, similar to the process of natural selection, only the best solutions remain. So, from the set of potential solutions to the problem competing with one another, the best solutions are chosen and combined with each other with the purpose of getting one universal solution from the set of solutions which will get better and better, similar to the process of organism population evolution. Genetic algorithms are used in data mining to formulate hypotheses about variable dependencies, in the shape of association rules or some other internal formalism. Careful selection of structure and parameters of the genetic algorithm can secure a significant chance to come up with globally optimal solution after an acceptable number of iterations. Genetic algorithms can be used for the process of decision making with the purpose of finding appropriate crops to grow which are profitable (Patcharanuntawat, Bhaktikul, Navanugraha, & Kongjun).

Nearest neighbor method is one of the most common techniques in data mining. Nearest neighbor method classifies an unknown sample based on known classification of its neighbors. Unlike the rest of the techniques, there is no learning process to create a model. Model is basically samples with known classification. Every sample needs to be classified similar to samples around it. 'Distance function plays a key role in the success of classification, as in many different data mining techniques' (Mucherino et al., 2009). All distances between an unknown sample and the other known samples can be calculated. Distance with the lowest value corresponds to the sample in the set of known samples which is closest to the unknown sample. So, unknown sample can be classified based on its nearest neighbor. However this rule of classification can be imprecise, because it is mostly based on one known sample. It can be precise if an unknown sample is surrounded by a few known samples with the same classification. If known samples have different classification, precision of this method declines. Essentially, if a high level of precision is required, all neighboring samples must be taken in consideration,

and unknown sample is classified based on the majority of known samples with the same classification. Weather variations have a large influence on crop yield. Recognizing the inherited variability of the weather, it is very desirable to evaluate potential methods of managing a number of probable weather sequences which directly influence the yield success. Research conducted (Rajagopalan and Lall, 1999) about application of nearest neighbor technique for predicting weather conditions showed that this technique is highly successful. Main point of the research is that this technique maintains cross dependency of the structure that generates rainfall independent of other variables. This is useful, for example, in acquiring appropriate attributes in comparison of snow and rain influence on yield success, where obtaining appropriate combination of meteorological variables makes a difference in the outcome of interest.

Application of Data Mining in Agriculture

Modern age has brought significant changes and information technologies in different areas of human activities have found wide application thus also in agriculture. Development and introduction of new information technologies which enable global networking, give agriculture the label of 'IT agriculture'. Information technologies increasingly provide assistance in systematic approach to solving agricultural problems. Access to the right information enables preparation of accurate reports, for example about using protective equipment, number of work hours of the machine on a specific crop, or a number of hired season work force. At the same time it is easier to keep track of work and verify exchange of information. Agriculture is abundant with diverse information which conditions the necessity to use data mining. Through data mining agricultural organizations are capable of producing descriptive and predictive information as support to decision making. Some authors (Buttle, 2009) claim that data from data mining can be used primarily for:

- Division of market and customers into segments,
- Identification of valuable clients and potential customers in the future,
- Investigation of causes in customer behavior,
- Defining different prices for individual customer segments,
- Identification of poor payers,
- Creating customer profiles that the organization desires to acquire and keep,
- Identify successful tactics for keeping and acquiring customers.

As is known agriculture is complex area where every day new knowledge is accumulated at increasing rate. Large portion of this knowledge is in the form of written documents, large part resulting from studies conducted on data and information acquired in agriculture from customers. Today there is a great tendency to make this information available in electronic format, converting information into knowledge, which is no easy task. With the increase in costs in agricultural enterprises and increasing necessity to control these costs, appropriate analysis of agricultural information has become the question of great importance. All agricultural enterprises need professional analysis of their agricultural data, undertaking which is long term and very expensive. Agricultural enterprises are very oriented on using information about market of agricultural products. Ability to use data in databases to extract useful information for high quality agriculture is a key to success of agricultural enterprises. Agricultural information systems contain massive amounts of information including information about crops, customers, market… With the use of data mining methods, useful patterns of information can be found in this information, which will be used for further research and report evaluation. Very important question is how to classify large amount of data. Automatic classification is done on the basis of similarities present in the data. This sort of classification is only useful if the conclusion acquired is acceptable for the agronomist or the end customer. Problem of predicting production yield can be solved with data mining techniques. It should be considered that the sensor data are available for some past tense, in which appropriate production yield was recorded. All this information creates a set of data which can be used for learning ways of classifying future production yields, because new sensor data are available. There are different techniques of data mining that can be used with this purpose (Mucherino and Ruß, 2011).

Yield prediction is a very significant problem of agricultural organizations. Each agriculturist wants as soon as possible to know how much yield to expect. Attempts to solve this problem date back to the time when first farmers started cultivating soil aiming to acquire harvest (Ruß and Brenning, 2010). Over the years, yield prediction was conducted based on farmer's experience of certain agricultural cultures and crops. However, this information can be acquired with the use of modern technologies, like GPS. Today acquiring large amount of sensor data is relatively quick, so agriculturists not only reap crops but also ever larger amounts of data. These data are processed and often interconnected (Ruß and Brenning, 2010).

In agricultural research, data mining begins with a hypothesis and the results are adjusted to fit that hypothesis.

This is different than the standard data mining practice, which simply begins with a set of data with no apparent hypothesis. While traditional data mining is directed to patterns and trends in datasets, agricultural data mining is more focused on the majority which stands out from patterns and trends. The fact that standard data mining is more focused on description and not on explanation of patterns and trends is what deepens the difference between standard and agricultural data mining.

Difficulties of Applying Data Mining in Agriculture

Development of ICT in agriculture has enabled the creation of electronic record about customers which are acquired by monitoring customers. These electronic records include a series of information: customer demographics, sales progress records, crop examination details, used protective equipment, previous crop rotation data, pedological findings. Information system automates and simplifies the work flow of agricultural enterprises. Many techniques were developed for learning about rules and relations automatically from datasets, for the purpose of simplifying tedious and erroneous knowledge gaining processes from empirical data. Since these techniques are credible, backed up by theory and adequate for the majority of data, they depend on their readiness to design real data (Veenadhari et al., 2011).

Integration of data mining in information systems of agricultural organizations reduces subjectivity in decision making and provides new useful agricultural knowledge. Predictive models provide the best support of knowledge and experience to agricultural workers. The problem of prediction in agriculture can be divided into two phases: learning phase and decision making phase. In learning phase, large dataset is transformed in a reduced (simplified) dataset. Number of characteristics and objects in this new set is much smaller than in the original set in more ways than one. Rules generated in this phase are used to make a more precise decision later. Newly formed dataset is used to generate predictions when new events with unknown outcomes appear with the prediction algorithm. This algorithm compares characteristics of the new object with the characteristics of objects represented in the chosen dataset. In case there is a match, new object is assigned an outcome which is equal to the corresponding object in the set. Goal of predictive data mining in agriculture is to build a predictive model which is clear, makes reliable predictions and assists agronomists to improve their prognoses, procedures for planning treatment of agricultural cultures. Important questions arise here which need answering: are the data and corresponding predictive characteristics enough to build a predictive model of acceptable performance; what is the relationship between attributes and outcomes; can an interesting combination or relationship be found between attributes; can certain immediate factors be extracted from original attributes which can increase performance of the predictive model.

Managerial tasks governing quality of agricultural services can be described as optimization of clinical processes in the sense of agricultural and administrative quality as well as cost/profit relationship. Key question in the process of management of quality of agriculture are quality of data, standards, plans and treatments.

Obstacles Faced by Data Mining in Agriculture

Numerous problems hinder the use of data mining in agriculture primarily when using ICT but the crucial problem in data mining in agriculture is that the raw data is vast and heterogeneous. These data can be acquired from different sources like sales records, field books, laboratory findings, opinions of professional agricultural services. All these components may have a large influence on diagnosis, prognosis and treatment of crops, and cannot be ignored if an agricultural organization wishes to achieve success. Scope and complexity of agricultural data is one of the obstacles for successful data mining. Missing, invalid, inconsistent or nonstandard data like parts of information recorded in different formats from different data sources create a large obstacle for successful data mining. It is very difficult for people to process gigabytes of records, although work with images is relatively easier, because they are capable of recognizing patterns, accept basic trends in data, and formulate rational decisions. Stored information is becoming less useful if they aren't in an easily understandable format. Agricultural data are often connected with uncertainty because of inaccuracy of assessment, sample variability, obsolete data. Over the last few years (Sharma and Mehta, 2012), numerous studies have been conducted on managing uncertain data in databases, like presenting uncertainty in a database/file and finding data which contain uncertainty. However, few studies have been conducted regarding problems of researching uncertain data. To apply traditional techniques of data mining, uncertain data should be reduced to atomic values. Inconsistencies in acquired and real values can significantly influence the quality of results in research (Sharma and Mehta, 2012).

The role of visualization techniques increases, because images are easiest for people to understand, and can provide a large amount of information in one image of results. Data mining in agriculture can be limited in access to data, since raw input for data mining often exist in different settings and systems, like administration, agricultural professional services, services of the ministry of agriculture, laboratories and other. Therefore, data must be acquired and integrated before starting data mining. Construction of data storage for data mining can be very expensive and time consuming process. Agricultural organizations developing data mining must use large investment resources, especially time, effort and money. Data mining project can fail for a number of reasons, like lack of support from management, inadequate data mining expertise, disinterest of agronomists and professional services.

Conclusion

Agricultural organizations and their management try every day to find information (knowledge) in large databases for business decision making. Often the case is that the solution for their problems was within their reach and the competition has already used this information. Data mining, through better management and data analysis, can assist agricultural organizations to achieve greater profit. Therefore it is crucial that managers of agricultural organizations get to learn about the idea and techniques of DM, because the amount of available information is sure to grow in the future, and it will not become clearer and easier to understand and make decisions. It is clear that the competition will not sit idly by, and ignore benefits these techniques can bring. Understanding of the processes which are carried out and decisions being made in agricultural organizations is enabled through data mining. By the use of data mining technique acquired knowledge can be used to make successful decisions which will advance the success of the agricultural organization on the market. Data mining requires a certain technology and analytic techniques, as well as the systems for reporting and monitoring which can measure results. Data mining once started, presents an endless cycle of acquiring knowledge. For organizations it represents on of the key points to create a business strategy on. Great efforts are invested in finding a more successful application of data mining in agricultural organizations. Primary potential lies in the possibility to study hidden patterns in datasets in agricultural domain. These patterns can be used for diagnosing crop condition, prognosing market development, monitoring customer solvency. However, available raw agricultural data are widely distributed, by nature different, and extensive. These data should be collected and stored in organized forms, and integrated to form an information system of an agricultural organization. Data mining technology provides user oriented access to new and hidden patterns in data, from which knowledge is generated which can help with decision making in agricultural organizations. Agricultural institutions use data mining technique and applications for different areas, for instance agronomists use patterns measuring growth indicators of plants, crop quality indicators, success of taken agrotechnical measures and managers of agricultural ogranizations pay attention on user satisfaction and economicaly optimal decisions. Use of data mining technique in agricultural organizations creates conditions for makind adequate decisions and with that achieving competing advantage.

References

Abdullah, A. and A. Hussain, 2006. Data Mining a New Pilot Agriculture Extension Data Warehouse. *Journal of Research and Practice in Information Tehnology*, **3** (3): 229-240.

boirefillergroup.com., 2010. Data Mining Methodology. Retrieved 06 12, 2012, from Boire-Filler Group: http://www.boirefillergroup.com/methodology.php

Buttle, F., 2009. Introduction to Customer Relationship Management. In: F. Buttle, Customer Relationship Management, Oxford: Butterworth-Heinemann is an imprint of Elsevier, pp. 1-25.

Cunningham, S. J. and G. Holmes, 1999. Developing Innovative Applications in Agriculture Using Data Mining. Department of Computer Science University of Waikato Hamilton, New Zealand.

Fayyad, U., G. P. Shapiro and P. Smyth, 1996. From Data Mining to Knowledge Discovery in Databases. American Association for Artificial Intelligence, pp. 37-54.

Foss, B. and M. Stone, 2001. Sucesful Custeomar Relationship Marketing. London: Kogan Page Limited.

Getoor, L., 2003. Link Mining: A New Data Mining Challenge. *SIGKDD Explorations*, **4** (2).

Huang, Y., Lan, Y., Thomson, S. J., Fang, A., Hoffmann, W. C. and R. E. Lacey, 2010. Development of soft computing and applications in agricultural and biological engineering. *Computers and Electronics in Agriculture*, **71:** 107–127.

Miller, D., J. McCarthy and A. Zakzeski, 2009. A Fresh Approach to Agricultural Statistics: Data Mining and Remote Sensing. *Section on Government Statistics – JSM 2009*, pp. 3144-3155.

Milovic, B., 2012. Kompleksnost implementacije koncepta CRM. *INFOTEH-JAHORINA,* **11:** 528-532.

Milovic, B., 2011. Upotreba Data Mininga u Donošenju Poslovnih Odluka. *YU Info 2012 & ICIST 2012,* pp. 153-157.

Mucherino, A. and G. Ruß, 2011. Recent Developments in Data Mining and Agriculture. Retrieved November 30, 2012, from http://www.antoniomucherino.it: http://www.antoniomucherino.it/download/papers/MucherinoRuss-dma11.pdf

Mucherino, A., P. J. Papajorgji and P. M. Pardalos, 2009. Data Mining in Agriculture. Springer Optimization and Its Applications, **34.**

Novkovic, N., Ž Ilin, M. Matković and B. Milovic, 2011. Informacione osnove za upravljanje proizvodnjom zdravstveno bezbednog povrća. Mediterenski Dani Trebinje 2011- Turizam i ruralni razvoj (pp. 373-379). Trebinje: Sajamski Grad Trebinje.

Patcharanuntawat, P., K. Bhaktikul, C. Navanugraha and T. Kongjun, 2007. Optimization for cash crop planning using genetic algorithm: A case study of upper Mun Basin, Nakhon Ratchasima province. 4th INWEPF Steering Meeting and Symposium, Paper 2-07: 2-13.

Rajagopalan, B. and U. Lall, 1999. A k–nearest-neighbor simulator for daily precipitation and other weather variables. *Water Resources Research*, **35** (10): 3089–3101.

Rajesh, D., 2011. Application of Spatial Data Mining for Agriculture. International Journal of Computer Applications, pp. 7-9.

Ramesh, D. and B. Varhadan Vishnu, 2013. Data Mining Techniques and Applicatins to Agricultural Yield Data. *International Journal of Advanced Research in Computer and Communications Engineering,* **2** (9): 3477-3482.

Ruß, G. and A. Brenning, 2010. Data Mining in Precision Agriculture: Management of Spatial Information. 13th International Conference on Information Processing and Management of Uncertainty, Proceedings (pp. 350-359). Dortmund, Germany: IPMU 2010.

Sharma, L. and N. Mehta, 2012. Data Mining Techniques: A Tool for Knowledge Management System in Agriculture. *International Journal of Scientific and Technology Research,* **1** (5): 67-73.

Štastny, J., V. Konečny and O. Trenz, 2011. Agricultural prediction by means of neural network. *Agric. Econ.* – Czech, **57** (7): 356–361.

Yang, Q. and X. Wu, 2006. 10 Challenging problems in data mining research. *International Journal of Information Technology & Decision Making,* **5** (4): 597–604.

Veenadhari, S., B. Misra and C. Singh, 2011. Data mining Techniques for Predicting Crop Productivity. *International Journal of Computer Science and Technology*, pp. 98-100.

Zaïane, O. R., 1990. Principles of Knowledge Discovery in Databases. Department of Computing Science, University of Alberta.

Received February, 9, 2014; accepted for printing December, 2, 2014.