Use of data mining methods in planning and management at Russian wood enterprises

I A Zakharenkova*, T P Belyaeva, I N Igotti and T O Terenteva

St. Petersburg State Forest Technical University, 5 Institutskiy Lane, St. Petersburg 194021, Russian Federation

*Corresponding e-mail: irina_hs@bk.ru

Abstract. The purpose of this article is to disclose utilization capacity of the Data Mining method for data mining (DM) at Russian wood enterprises. The authors provide a rationale for use cases of the data mining technology – a combination of mathematical tools and recent developments in the technology industry – for the purpose of efficient system planning and management. On the basis of researches previously conducted by the authors, a list of problems and tasks is provided by the areas of activities of wood enterprises, resolved through the use of Data Mining methods. DM significance and relevance have been emphasized with regard to adoption of efficient managerial decisions in the area of production at Russian wood enterprises.

1. Introduction
Rapid development of information technologies, progress in methods of data collection, storage and processing allowed many organizations to collect enormous arrays of information, which are required to be analyzed.

In July 2020 the Federal Agency for Forestry (Rosleshoz) announced creation of a single IT-platform for interaction between private and public sectors. It will become part of the all-Russian digital resource, which will be created at the Rosleshoz site. The system will help to transfer interaction between the timber private and public sectors to electronic format. This initiative is critical for the timber industry (TI) in general. However, at present the problem of digital technology application in diagnostics and implementation of various decisions at the level of specific branch enterprises is of no small importance.

Despite a wide variety of digital technologies for the TI in general and wood enterprises, at present there are comparatively few real-world examples of their actual usage, as far as hardly anyone understands, how digital technologies pass into business processes.

2. Experimental Part
At present, a tendency associated with intellectualization of methods of data management and analysis is actively developing in the world. The data mining systems are aimed to minimize efforts of managers and supervisors at different levels in the data analysis process, and to adjust analysis algorithms. Many data mining systems allow not only to resolve classical problems of decision-making, but are also able to identify cause-and-effect relations, latent patterns. Methods of generation of such models usually fall within the domain of artificial intelligence. One of such methods is the Data Mining system.
English term “Data Mining” has no precise translation into Russian (data dissection, data/information extraction), therefore in most cases it is used in the original. The term “Data mining” is considered to be the most appropriate indirect translation.

Data mining divided into problems of classification, modelling, forecasting and others, represents identification of latent patterns or relations between variables in large-scale arrays of raw data [1].

“Data Mining” term was introduced by Gregory Piatetsky-Shapiro—author of books and collected works on data mining in 1989.

“Knowledge Discovery in Databases” (KDD) term is applied as a synonym of Data Mining [2].

The purpose of Data Mining methods is to discover previously unknown, unconventional, practically useful knowledge. DM lies at the intersection of such fields as artificial intelligence (neural networks and computer-assisted learning), mathematical statistics and database theory.

As far as Data Mining is across-disciplinary area, a lot of methods and algorithms arise, which are implemented in different current Data Mining systems. Many of such systems integrate several approaches at once. Nevertheless, each system has as a rule a key component, on which the focus is made (see figure 1).

**Figure 1.** Areas of Data Mining application.

Data Mining is a process of discovery of raw data previously unknown, unconventional, useful from a practical standpoint, available for interpretation and required to make decisions in different fields of human activity [1].

DM model generation is part of a large-scale process, beginning with formulation of questions on data and development of a model to answer such questions before the model deployment in the production environment.

This process may be presented as a sequence of the following six basic steps (see figure 2).

**Figure 2.** Process of Data Mining (DM) with the use of Data Mining methods.

1. Statement of the problem.
2. Data acquisition.
3. Data examination.
4. Model generation.
5. Model examination and evaluation.
6. Obtainment of the final model.

Data Mining methods and algorithms include:
- artificial neural networks;
- decision trees, character rules;
- a support vector machine;
- Bayesian networks (graphs probable models);
- linear regression;
- correlation and regression analysis;
- non-hierarchical clustering methods;
- evolutionary programming and genetic algorithms;
- a limited search method, various visualization techniques, etc.

Among analytical methods used in the Data Mining technology, known mathematical algorithms are applied. The new thing in application thereof is their applicability upon resolution of specific problems conditioned by emerging functionalities of hardware and software products. Moreover, many Data Mining methods have been developed within the framework of the artificial intelligence theory.

DM represents a process of discovery of usable parameters in Big Data. Mathematical analysis is applied in DM to identify patterns and tendencies. Usually, it is impossible to identify such patterns upon traditional data viewing due to overcomplicated Big Data associations.

Such patterns and tendencies may be put together and defined as a DM model. DM models may be applied to particular scenarios of development of any wood enterprise considered in the context of the Data Mining application.

Primary areas of the Data Mining technology application at timber industry manufacturing enterprises are as follows.
1. Forecasting: evaluation of sales, forecasting of the company profitability and efficiency levels.
2. Risk and probability: identification of an equilibrium point for risky scenarios of the company development strategy implementation.
3. Recommendations: identification of product types, which may be efficiently sold at particular markets.
4. Search for sequences: supplier selection analysis when purchasing resources, forecasting of the next potential event in general diagnostics of the situation.
5. Grouping: making groups by related elements, analysis and forecasting of the key factors, which affect an event or an indicator.

Traditional data analysis methods (statistical methods) and Online Analytical Processing (OLAP), customary for practice of using by TI enterprises, are basically focused on verification of previously formulated hypotheses, while Data Mining is focused on searching for unobvious patterns. Data Mining tools may find such patterns and independently form hypotheses on interconnections. However, OLAP is more suitable for historical data analysis (subject to variations with time), while Data Mining relies on historical data to receive answers to any questions concerning the future, which is particularly true for decision-making in the context of continually changing external environment.

Data Mining capabilities are practically unlimited for application in practical forecasting of performance targets of modern TI manufacturing companies, generation of scenarios of development in the long-term.

Previously we have published a methodology for developing a model of dependence of key company performance indicators – sales profit and unit costs on different factors, which operate in competitive environment, as exemplified by a modern Russian furniture company, and for determining the nature of correlations and the degree of factor influence with the use of Data Mining methods [3]. The methodology is based on studying correlations between key factor features by means...
of pair correlation linear coefficient analysis based on statistical evaluation of input information on the company performance indicators for a number of years. Further, based on the selected model, a forecast of unit costs and sales profit of the company is determined [4].

The use of methods of Data Mining correlation and regression analysis, which take into consideration the factors of external and internal environment of the company in this methodology, allows deriving universal established models for calculation of sales profit and unit costs for the purpose of using them for development of scenarios for making planned decisions by the furniture company, which proved the adaptive nature of application of the Data Mining tools in activities of Russian wood enterprises.

3. Results and Discussion

Industry characteristics of industrial production and technical processes at TI enterprises in Russia provide good opportunities for using the Data Mining technology in the course of solution of different problems associated with enhancement of performance efficiency and system planning at wood enterprises.

As far as one of the critical parameters of technical and production processes at TI enterprises is their controllability with deviations traced, which generally lie within the boundaries known in advance and are relatively stable, the possibilities of application of the Data Mining technology, which handles weakly structured problems with regard to organization of production, are hardly of current interest, however, they may be successfully used in diagnostics of environmental parameters, as well as parameters associated with such production activity indicators as product quality, operative planning of the production process, diagnostics of disproportions and interruptions of production processes, generation of long-term scenarios of production activity.

It should be also noted that one of the most important purposes of the Data Mining methods is visualization of computing results, which allows the Data Mining architecture and tools to be used by people, who have no special mathematical training. This is especially relevant to an average TI enterprise, which is short-staffed, as a rule, with specialists, who possess competences in statistical methods of data analysis, theory of probability and mathematical statistics.

The principal area of the Data Mining technology utilization in practical operation of a wood enterprise today is making managerial decisions of different levels of complexity when planning under uncertainty, associated with achievement of strategic objectives of the company development in the long-term.

Wide implications of DM in practical management and planning are also associated with development of scientifically founded forecasts. In which case a time horizon, with regard to which the forecasts are made, is of no importance. Analysis of the company development prospects through appraisal of scientifically founded concepts of the future development is a wide scope of work to use the Data Mining technology.

Project and program forecasting, as well as development of exploratory and normative forecasts may be implemented with the use of graphs probable models, correlation and regression analysis, limited search methods through the use of Data Mining techniques.

Primary areas may be pointed out, within the framework of which the Data Mining methods may be used in respect to organization and planning of general production and economic activities of TI enterprises (see table 1).

Areas of activities of wood enterprises for each DM task may be extended. Essentially, any area of activities of a wood enterprise or a group of enterprises, which contains Big Data bases for the purpose of processing information and compiling results by a specific algorithm, may serve as a field of operation for DM techniques application.
Table 1. Application of the Data Mining technology options to make optimal planned and managerial decisions at wood enterprises.

| Primary objectives of Data Mining for wood enterprises | Area of activity                      | % of problems solved |
|-------------------------------------------------------|---------------------------------------|----------------------|
| 1. System analysis of production processes             | Planning, system management           | up to 70%           |
| 2. Planning and forecasting of production management development at enterprises in the short-term and the long-term | Planning, system management           | up to 80%           |
| 3. Optimization of planned decisions within the framework of the assumed activity planning model | Planning, system management           | 70-90%              |
| 4. Generation of forecasts and scenarios of product quality changes depending on key parameters | R&D, innovations, product quality monitoring | 70-90%              |
| 5. Search for latent tendencies (impact factors and uncertainty factors), which affect production organization and management processes | R&D, production                      | 70-80%              |
| 6. Diagnostics of patterns in development of production situations associated with an uncertainty factor | Production                           | 60-80%              |
| 7. Forecasting of change matrices in relation to stable patterns in development of the industry production processes | R&D, production, system management    | 80-90%              |
| 8. Diagnostics and substantiation of the degree of dependency between the production process parameters and the key impact factors | R&D, production                      | 60-80%              |
| 9. Identification of internal characteristics of the production process environment and generation of scenarios of its alteration to specify planned performance indicators | Planning, forecasting, production     | 80-100%             |
| 10. Optimization of key recommendations on management of the output production processes within a limited planning horizon | Planning, production                 | 80-100%             |
| 11. Drafting and visualization of results of forecasting production parameter changes depending on the key uncertainty factors | Forecasting, logistics, production    | 80-90%              |
| 12. Reporting on options of searching for optimal planned and managerial decisions with a full case of reliable, acceptable and the most efficient decisions | Planning, system management           | 90-100%             |
| 13. Segmentation of customers based on their profitability, CRM (Customer Relationship Management) | Marketing, sales                      | 70-90%              |
| 14. Ranking key factors, which affect key indicators of profitability, and evaluation of the degree of impact within the framework of far-term models of strategic planning | Planning, finances                   | up to 100%          |

Please note that the percentage of problems solved at wood enterprises through the use of Data Mining architecture and technologies is conventional to a certain extent, however, it approximates to evaluation of relevance of the Data Mining techniques usage in operation of modern international companies [1].

Areas of activities of TI enterprises, where you can successfully use Data Mining tools, are determined by the degree of complexity and importance of problems solved within the context of adoption of managerial and planned decisions or Big Data diagnostics and analysis.
4. Conclusion
Applicability of the Data Mining methods and techniques in practical operation of modern wood enterprises is explained by added complexity of searching for optimal decisions, a need for extensive analysis and diagnostics of the performance results in the context of the Internet economy (digital economy), when a business tendency is formed for searching for development trends with regard to markets and systems, planning and forecasting of the company development.

Therefore, the use of the Data Mining methods by domestic wood enterprises gives distinct advantages and facilitates unification of program and managerial decisions in the context of development of e-commerce and Internet media. Moreover, in the absence of any consistency in formats of data interchange, analytical processing, modelling and visualization of industry data, the Data Mining methods are a tool for planning and development of TI enterprises, obtainment by them of distinct competitive advantages.

References
[1] Barseghyan A A, Kupriyanov M S, Stepanenko V V and Holod I I 2004 Data mining methods and models: OLAP and Data Mining. Education guidance, 2nd edition, (Saint-Petersburg: BHV-Petersburg) p 336
[2] Matthew R 2020 Data mining. Data mining Facebook, Twitter, LinkedIn, Instagram, GitHub. (Saint-Petersburg: PETER) p 464
[3] Zakharenkova I A, Belyaeva T P and Terenteva T O 2020 Use of Data Mining to optimize the selection of scenarios for accepting planned decisions in Russian companies Proceedings of the 35th International Business Information Management Association (IBIMA) conference, 1-2 April 2020, Seville, Spain
[4] Zakharenkova I A and Igotti I N 2018 Theoretical aspects of using models upon managerial decision-making for enterprise performance improvement Scientific forum: Economics and management: collection of essays following the 13th international research to practice conference (Moscow: Publishing house MCCME) 1 (13) pp 46-51