

# Collaborative Predictive Business Intelligence Model for Spare Parts Inventory Replenishment

Nenad Stefanovic

Faculty of Technical Sciences, Svetog Save 65,  
32000 Cacak, Serbia  
nenads@kg.ac.rs

**Abstract.** In today's volatile and turbulent business environment, supply chains face great challenges when making supply and demand decisions. Making optimal inventory replenishment decision became critical for successful supply chain management. Existing traditional inventory management approaches and technologies showed as inadequate for these tasks. Current business environment requires new methods that incorporate more intelligent technologies and tools capable to make fast, accurate and reliable predictions. This paper deals with data mining applications for the supply chain inventory management. It describes the unified business intelligence semantic model, coupled with a data warehouse to employ data mining technology to provide accurate and up-to-date information for better inventory management decisions and to deliver this information to relevant decision makers in a user-friendly manner. Experiments carried out with the real data set, from the automotive industry, showed very good accuracy and performance of the model which makes it suitable for collaborative and more informed inventory decision making.

**Keywords:** predictive analytics, supply chain inventory management, data mining, collaborative business intelligence, web portal.

## 1. Introduction

The success of many organizations depends on their ability to manage the flow of materials, information, and money into, within, and out of the organization. Such a flow is referred to as a supply chain [1]. Since supply chains are typically dispersed and can be very complex, involving many distinct organizations, there are numerous problems and challenges in their operations. These problems can cause high inventory costs, mismatch of supply and demand, lower customer satisfaction level, lost sales, decreased adaptability, etc.

The aim of the integrated supply chain planning and operations management is to combine and evaluate, from a systemic perspective, the decisions made and the actions undertaken within the various processes which compose the supply chain.

The need to optimize the supply chain, and therefore to have models and computerized tools for medium-term inventory planning and replenishment, is particularly critical in the face of the high complexity of current supply chain systems, which operate in a dynamic, uncertain and truly competitive environment.

It is not enough to know only what happened and what is happening now, but also what will happen in the future and how/why did something happen. Due to the complex interactions occurring between the different components of a supply chain, traditional methods and tools intended to support the inventory management activities seem today inadequate. Thus, predictive analytics and data mining became indispensable and valuable tools for making more intelligent decisions.

On the other hand, predictive models, themselves, are not enough. Information and knowledge derived from these analytical models needs to be delivered to all parties involved in supply chain inventory management. As a result, collaboration through dedicated web-based workspaces became essential for more efficient and effective coordination and decision making [2]. Furthermore, today's users require self-service business intelligence (BI) capabilities which allows them to perform advanced data analysis tasks without support of the specialized IT staff.

This paper presents the unified business intelligence semantic model for predictive inventory management which provides better collaboration, flexibility, richness, and scalability comparing to traditional BI models. It describes corresponding data mining models for making out-of-stock prediction of the automotive spare parts. The models are designed on top of the data warehouse which is loaded with sales data from the retail spare parts stores. Accuracy of the models is demonstrated through testing and evaluation of the results. Finally, the specialized analytical web portal which provides collaborative, self-service, personalized and secure analytical services is presented.

## 2. Background Research

Inventory control is the activity which organizes the availability of items to the customers. It coordinates the purchasing, manufacturing and distribution functions to meet marketing needs.

Inventory management is one of the most important segment of the supply chain management. Companies face the common challenge of ensuring adequate product/item stock levels across a number of inventory points throughout the supply chain. Additionally, uncertainty of demand, lead time and production schedule, and also the demand information distortion known as the bullwhip effect [3], makes it even more difficult to plan and manage inventories.

The basis for decision making should be information about customer demand. Demand information directly influence inventory control, production scheduling, and distribution plans of individual companies in the supply chain [4]. Making decision based on local data leads to inaccurate forecasts, excessive inventory, and less capacity utilization.

Generally, determining the adequate stock levels balances the following competing costs:

- Overstocking costs – these include costs for holding the safety stocks, for occupying additional storage space and transportation.
- Costs of lost sales – these are costs when customer wants to buy a product that is not available in that moment.

The best way to deal with these competing costs is to use data mining techniques to ensure that each inventory point (internal warehouse, work-in-process, distribution center, retail store) has the optimal stock levels.

Commonly, managers have relied on a combination of ERP (Enterprise Resource Planning), supply chain, and other specialized software packages, as well as their intuition to forecast inventory. However, in today's high uncertain environment and large quantities of disparate data demands new approaches for forecasting inventory across the entire chain. Data mining tools can be used to more accurately forecast particular product to the right location.

Although most data mining techniques existed for decades, it is only in the last years that commercial data mining gained wider acceptance. This is primarily due to the following factors [5]:

- There is a huge amount of data available
- The data is stored in data warehouses
- There are commercial data mining software
- Increasing competition and the growing interest of the business users

Forecasting and planning for inventory management has received considerable attention from the scientific community over the last 50 years because of its implications for decision making, both at the strategic level of an organization and at the operational level. Many influential contributions have been made in this area, reflecting different perspectives that have evolved in divergent strands of the literature, namely: system dynamics, control theory and forecasting theory [6].

Inventory management is probably the key supply chain process, and inventory costs represent a large portion of the total supply chain costs. Incorporating predictive analytics in an inventory management process can lead to many benefits such as cost reduction, higher customer service level, optimal reordering policy, enhanced productivity, shorter cash-to-cash cycle time, and ultimately increased profitability [7].

A number of research projects have demonstrated that the efficiency of inventory systems does not relate directly to demand forecasting performance, as measured by standard forecasting accuracy measures. When a forecasting method is used as an input to an inventory system, it should therefore always be evaluated with respect to its consequences for stock control through accuracy implications metrics, in addition to its performance on the standard accuracy measures [8].

Chandra and Gabris used simulation modeling to investigate the impact of the forecasting method selection on the bullwhip effect and inventory performance for the most downstream supply chain unit [9]. The study showed that application of autoregressive models compares favorably to other forecasting methods considered according to both the bullwhip effect and inventory performance criteria.

Liang and Huang employed multi-agents to simulate a supply chain [10]. Agents are coordinated to control inventory and minimize the total cost of a supply chain by sharing information and forecasting knowledge. The demand is forecasted with a genetic algorithm (GA). The results show that total cost decreases and the ordering variation curve becomes smooth.

Spare parts are very common in many industries and forecasting their requirements is an important operational issue. In recent years, there have been advances in forecasting methods for spare parts, demand information sharing strategies and the design of forecast support systems. Boylan and Syntetos give thorough review on these developments and provides avenues for further research are explored [11].

Accurate demand forecasting is of vital importance in inventory management of spare parts in process industries, while the intermittent nature makes demand forecasting for spare parts especially difficult. Hua et al. proposed an approach that provides a mechanism to integrate the demand autocorrelated process and the relationship between explanatory variables and the nonzero demand of spare parts during forecasting occurrences of nonzero demands over lead times [12]. The results show that this method produces more accurate forecasts of lead time demands than do exponential smoothing, Croston's method and Markov bootstrapping method.

Bala proposed an inventory forecasting model which use of purchase driven information instead of customers' demographic profile or other personal data for developing the decision tree for forecasting [13]. The methodology combines neural networks, ARIMA and decision trees.

Dhond et al [14] used neural-network based techniques for the inventory optimization in a medical distribution network which resulted in 50% lower stock levels. Symeonidis et al. [15] applied data mining technology in combination with the autonomous agent to forecast the price of the winning bid in a given order.

When it comes to supply chain performance management, there have not been many research efforts related to predictive performance measurement. Seifert and Eschenbaecher introduced a predictive performance measurement approach as a planning tool for virtual organizations to anticipate the performance of a planned virtual team [16]. Derrouiche et al. proposed an integrated framework for supply chain performance evaluation based on data mining techniques [17]. This framework enables the development of a predictive collaborative performance evolution model and decision making which has forward-looking collaborative capabilities. Stefanovic [18] presented the proactive supply chain performance management model which unifies business intelligence tools like data warehousing and data mining to integrate data and to perform predictions of the key performance indicators (KPI), and finally to deliver derived knowledge to decision makers via collaborative web portal. The results show that these models give very accurate KPI projections and provide valuable insights into newly emerging trends, opportunities, and problems.

Even though, forecasting is seen as a crucial segment of effective inventory management and supply, there are no many reports from the industry which demonstrate successful application of the prediction models and solutions. This is especially true when it comes to automotive spare parts supply management that is characterized by high uncertainty of demand and thousands of different parts. Most of the existing research is focused on specific segments of the analytical solutions (i.e. only predictions), without integral approach which comprises other crucial elements of the successful BI solution, such as data extraction, transformation, and loading, performance, dimensional modeling, collaboration, and information delivery [19].

Besides designing accurate and useful predictive data mining models, there is an important issue related to operational performance of these models. Today, in organizations, a large amount of data (also called, Big Data) is generated every day, thus making the data integration, storage and processing extremely challenging. This is especially true for supply chain BI systems, considering a multitude of partners, products, and complexity of the business processes. The main challenge is how to use huge amount of data and turn it into process improvement actions. Organizations still companies lack a clear roadmap for how to implement big data analytics in a meaningful and cost-effective manner. Sanders provides a systematic framework for

organizations on how to implement big data analytics across the supply chain to turn information into intelligence and achieve a competitive advantage [20]. Data science, predictive analytics and big data will transform the way supply chain are designed and managed, presenting a new and significant challenge to logistics and supply chain management (SCM)[21].

Another crucial aspect of successful predictive BI systems is related to collaborative decision making. One of the main emerging trends is toward the collaborative BI which bring together BI tools with collaboration tools and technologies such as enterprise content management (ECM), rich internet applications (RIA), wikis, and social networking. This synergy, called BI 2.0, is aimed to improve decision making process through the use of Web 2.0 technologies in order to facilitate exchange of knowledge and ideas [22]. Although, BI 2.0 concept is yet to be achieved by majority of organizations, there is an emergence of the BI 3.0 systems which should bring even better collaboration capabilities, mobile support, and automated decision making [23]. This should enable organization to utilize the collective intelligence, and also better coordination and alignment during the data-driven decision making process.

Even though, very promising approach, collaborative BI has not been researched sufficiently. Collaborative BI is aimed to extend the decision-making process beyond a single organization to entire supply chain. Most of the current BI solutions fail to meet the challenges of ad-hoc and collaborative decision support. There is a need for scalable and flexible platform for collaborative, ad-hoc BI over large data sets. This can be achieved by developing methodologies, concepts and an infrastructure to enable self-service analytics for business users and collaborative decision making [24]. Rizzi analyses several approaches for BI collaboration and proposes a peer-to-peer BI framework, where peers expose querying functionalities aimed at sharing business information for the decision-making process [25]. The main features of this framework are decentralization, scalability, and full autonomy of peers.

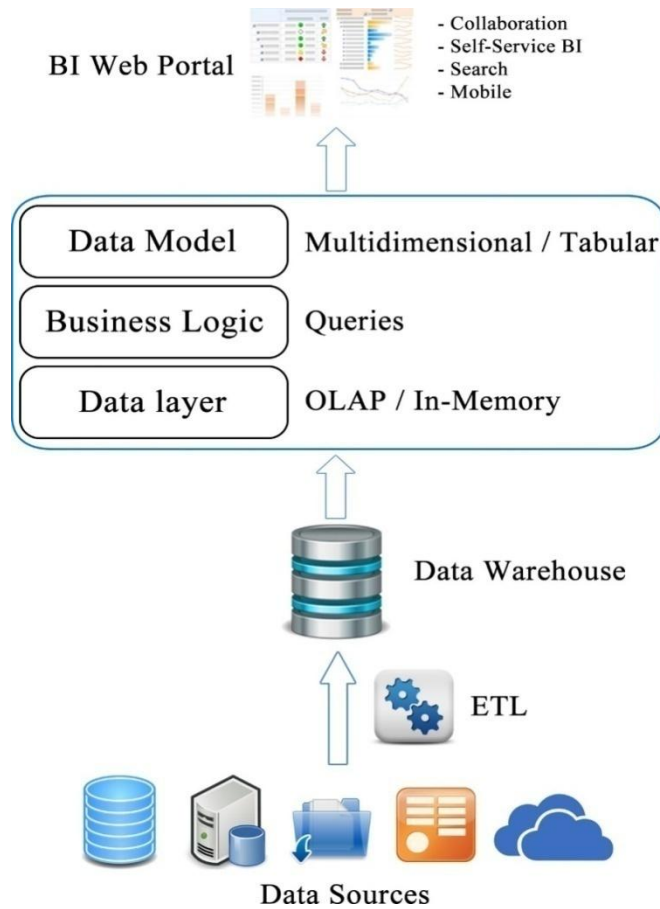
This last aspect of predictive analytics is probably the most important, because it is in this context that any real value is derived. It relates to taking actions based on the intelligence obtained from the BI systems [26]. There is a little value in the discovered knowledge if there are no appropriate action which lead to business process improvement. For example, the resulting information and knowledge can be used with special expert systems which can provide concrete actions based on the best practices contained in the knowledge bases [27].

This background research shows that although BI predictive applications have great potential in various business domains, there are still many challenges that should be overcome by incorporating innovative BI models and solutions. In the following sections, the unified BI semantic model that enable more collaborative, flexible and scalable BI solutions, with richer analytical capabilities is presented. It is the basis for various analytical applications such as forecasting, associations, segmentation, performance measurement, advanced reporting, etc. This BI model is used for designing the inventory out-of-stock predictive model which utilizes certain data mining algorithms. Data mining models and analytical cubes are designed in such a way to accommodate specificity of the automotive spare parts management in terms of products, stores, inventory, and time dimensions.

### 3. Collaborative BI Semantic Model

Today’s business environment, in which supply chain operate, requires a new and comprehensive approach to business intelligence systems that is more pervasive, flexible and user friendly. Most of the existing approaches focus on single aspect of the BI solution, neglecting important BI elements such as performance, information delivery, collaboration and process improvement.

In this paper, the collaborative BI semantic model that encompasses the entire BI lifecycle is presented (Figure 1).



**Fig. 1.** Collaborative Business Intelligence Semantic Model

The main advantages of the proposed model are:

- Integration – data from various data sources is first consolidated through the ETL (Extract, Transform, Load) process, into the single data warehouse, thus providing the basis for building analytical models.

- Flexibility – analytical models can be built in both, traditional multidimensional OLAP and in-memory tabular modes, depending on the business requirements and priorities.
- Richness – Semantic model enables business logic to be encapsulated by using appropriate analytical queries. Also, advanced construct can be added, such as, data mining models, KPIs (Key Performance Indicators), and actions.
- Performance – architecture of the model provides scalability and the performance is improved by using the in-memory data store.
- Collaboration – introduction of the specialized BI web portal with various collaborative and analytical services enables more pervasive BI, cooperative decision making, knowledge exchange, and process improvement.
- Usefulness – the proposed approach to design of BI systems supports different analytical scenarios, like personal (self-service), team, organizational and supply chain-wide business intelligence.

#### **4. Predictive Model for Inventory Management**

Automotive industry, as one of the most complex, faces considerable challenges. Shorter time-to-market, reduced product lifecycle, built-to-order strategies, pull systems, demand uncertainty, as also multitude of parties involved, forces companies to adopt new ways of doing business. SCM offers companies a new way to rapidly plan, organize, manage, measure, and deliver new products or services.

This section describes the business intelligence solution for the real automotive supply chain, which utilizes data warehouse and data mining technology to provide timely information for spare parts inventory management decisions. The presented methodology is designed to provide out-of-stock predictions at the location/product level. For a particular product, data mining model is built that makes out-of-stock predictions for each store in the chain. This approach enables more effective balance between the competing costs related with stocking.

##### **4.1. Data Warehouse Design**

In order to gather data from many distributed sources, we needed to extract, clean, transform and load data into the data warehouse that summarize sales data from 36 retail stores and for more than three thousands of different spare parts. These data are distributed among multiple heterogeneous data sources and in different formats (relational databases, spreadsheets, flat files and web services).

SQL Server Integration Services were used for the ETL, while SQL Server Analysis Service were used for data warehousing (cubes, dimensions, measures, etc.) and data mining. For bridging the gap between the user/developer and the data sources, the BI semantic model is used [28].

A BI semantic model is constructed over many physical data sources, allowing users to issue queries against the model, using one of a variety of client tools and programming technologies. The main advantages are a simpler, more readily understood

model of the data, isolation from heterogeneous backend data sources, and improved performance for summary type queries.

The following data sets are used for the out-of-stock predictive modeling:

- Sales data that is aggregated at the store, product (part), and day level. Daily sales are stored for each product that is sold, for each store in the retailer's chain.
- Inventory data that is aggregated at the store, product (part), and day level. This is the number of days that the product has been in stock, for each product, for each day, and for each store.
- Product (part) information such as product code, name, description, price, and product category.
- Store information such as store description, store classification, store division, store region, store district, city, zip code, space capacity, and other store information.
- Date information that maps fact-level date identifiers to appropriate fiscal weeks, months, quarters, and years.

The data warehouse is the basis for all business intelligence applications and particularly for data mining tasks. Data warehouse allows us to define better data mining models based on the constructed data warehouse to discover trends and predict outcomes.

#### 4.2. Data Mining Methodology

In order to increase the quality and accuracy of the forecasts, we have applied a two-phase modeling process. Phase I of the modeling process consists of clustering stores in the supply chain based upon aggregate sales patterns. After store-cluster models have been constructed, in phase II, these clusters are used to more accurately make out-of-stock predictions at the store/product level [29].

The general data mining process is shown in Figure 2. The process begins analyzing the data, choosing the right algorithm in order to build the model. The next step is model training over the sampled data. After that, model is tested, and if satisfactory, the prediction is performed.

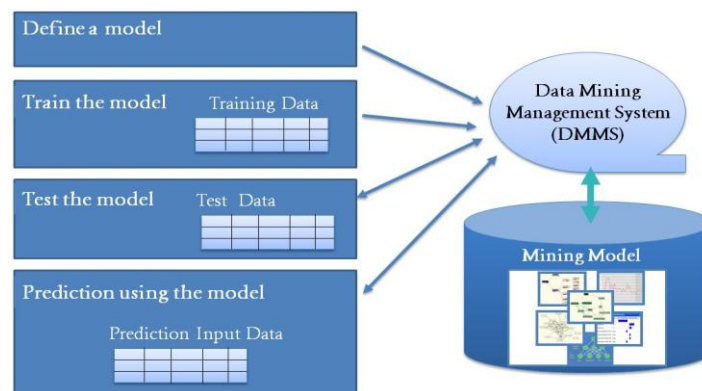


Fig. 2. Data mining process



### 4.3. Inventory Predictive Modeling Process

Phase I consists of grouping together those stores that have similar aggregate sales patterns across the chain. Store clustering is accomplished by using the data mining Clustering algorithm. Dataset holds aggregate sales patterns and Clustering algorithm groups together stores into clusters. The modeling dataset is based on aggregate sales data that is derived from the data warehouse. The measure that is used to group together stores is computed over this aggregate sales data.

In phase II, cluster models were used to build more accurate out-of-stock forecasting models. This allows predictive algorithms such as Decision Trees and neural Networks to use the results of the clustering process to improve forecasting quality. In essence, to make the predictions for a given spare part  $p$  in a given store  $s$ , the forecasting algorithms use the fact that the sales for the same spare part  $p$  in a similar store  $s$  may produce better results when determining whether or not a particular part will be out of stock in a particular store.

Modeling process consists of the following high-level steps:

1. Use the spare part hierarchy in the product information (dimension) portion of the data warehouse to determine the spare part category  $c(p)$  for part  $p$ . We assume that spare parts within the same category have similar aggregate sales patterns across the chain of stores, and the product hierarchy is used to identify the set of similar products  $c(p)$  for a given product  $p$ . Alternatively, a product clustering approach could be used to determine a data-driven grouping of spare parts similar to  $p$  by clustering parts based upon their sales across the chain of stores.
2. Prepare modeling dataset  $D_{\text{cluster}}$  for store clustering to capture store-level properties and sales for category  $c(p)$ .
3. Apply the Clustering algorithm to the dataset  $D_{\text{cluster}}$  to obtain  $k$  clusters (groups) of those stores that are similar across store-level properties and sales for category  $c(p)$ .
4. For each cluster  $l = 1, \dots, k$  obtained in previous step:
  - a. Let  $S(l)$  be the set of stores that belong to cluster  $l$ . These stores have similar category-level aggregate sales, for the category  $c(p)$ .
  - b. Create a dataset  $D_{\text{inventory}}(p, S(l))$  consisting of historic and current weekly sales aggregates, and changes in weekly sales aggregates, for each store  $s$  in  $S(l)$ . In addition, include Boolean flags indicating whether or not product  $p$  was in stock or out of stock one week into the future and two weeks into the future.
  - c. Apply the predictive modeling algorithms (in this case Decision Trees and Neural Networks) to the dataset  $D_{\text{inventory}}(p, S(l))$ . Use the historic and current weekly sales aggregates as input attributes and the one- and two-week out-of-stock Boolean flags as output or predict-only attributes. This instructs data mining engine to generate a model that takes as its input the historic and current weekly sales, along with changes in weekly sales, and then make a prediction of the Boolean flags that indicate whether or not spare part  $p$  will be out of stock one and two weeks into the future.

**Phase I: Store clustering.** The goal of store clustering is to obtain groups of stores that have similar sales patterns, focused on sales over the spare parts in the category to which part  $p$  belongs  $c(p)$ . Phase I begins with constructing the dataset that will be used for store clustering.

The dataset used for store clustering consisted of store-level aggregate sales over the time period of four years. Typically, the dataset consists of a single table with the unique key (StoreID) that identifies each item (store in the chain). The creation of this table can be automated by designing the appropriate ETL package. However, we decided to take advantage of the BI semantic model, and defined the data source view against it. This way, denormalized data source view is created over normalized set of fact and dimension data and without worrying about underlying data sources.

The store clustering task is to group together stores based upon similarity of aggregate sales patterns. Firstly, we had to identify a set of aggregate sales attributes relevant for this project. Attributes were aggregated over the fact data in the data warehouse. These attributes are category-specific (*total\_sale\_quantity*, *total\_sale\_amount*, *quantity\_on\_order*, *discount\_amount*, etc.) and store-specific (*total\_sales*, *total\_weekly\_on\_hand*, *total\_weekly\_on\_order*, etc.).

After initial business understanding phase, data cleaning and transformation, data warehouse construction and loading, the next step is clustering mining model construction.

Cases (i.e. stores) within the same group have more or less similar attribute values. The mining structure defines the column structure that will be used to construct the store-clustering model. All attributes are selected as *input* attributes except the *Category\_Fraction\_Sales* (fraction of total non-discount sales coming from parts in category  $c(p)$  in the given store) and *Category\_Total\_Sales\_Quantity* (total quantity of spare parts in category  $c(p)$  that were sold during the non-discount period) attributes that are selected as *predict*.

Two clustering algorithm parameters were tuned in order to get better outcome. *Cluster\_Count* parameter specifies the maximum number of clusters to search for in the source data. In order to produce distinct clusters that sufficiently capture the correlations in store properties and aggregate sales/inventory values, the *Cluster\_Count* parameter was altered and tested with different values to obtain desired results. The other parameter *Minimum\_Support* instruct clustering algorithm to identify only those clusters that have given value or more cases (stores in our case) in them. After setting the parameters for the Clustering algorithm, the mining structure is processed, thereby creating and populating the mining model. Figure 2 shows store clustering mining structure and algorithm parameters.

In the model-building stage, we build a set of models using different algorithms and parameter settings. After the store-clustering models have been constructed, they are evaluated by using the Cluster Browser to determine if the clusters are distinguished by category sales patterns.

The store clusters tend to be discriminated primarily by the *total\_sales*, *category\_sales\_quantity*, *category\_weekly\_sales*, *category\_weekly\_on-hand*, and *on-order* values. Figure 4 shows the derived store clusters shaded with the different density consistent with the population values, and also the link density relationships.

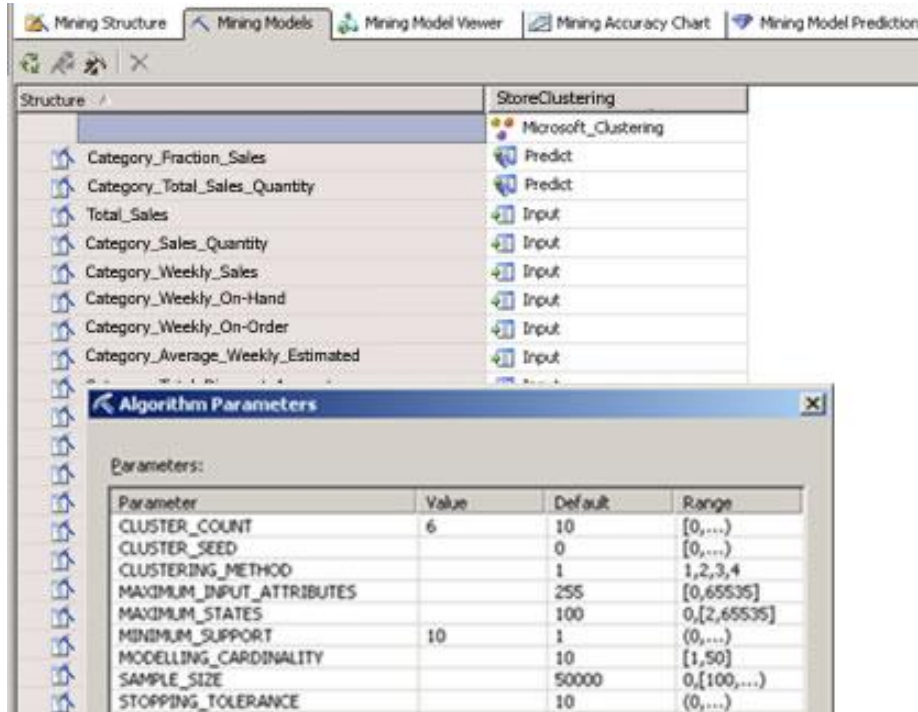


Fig. 3. Store-clustering mining structure

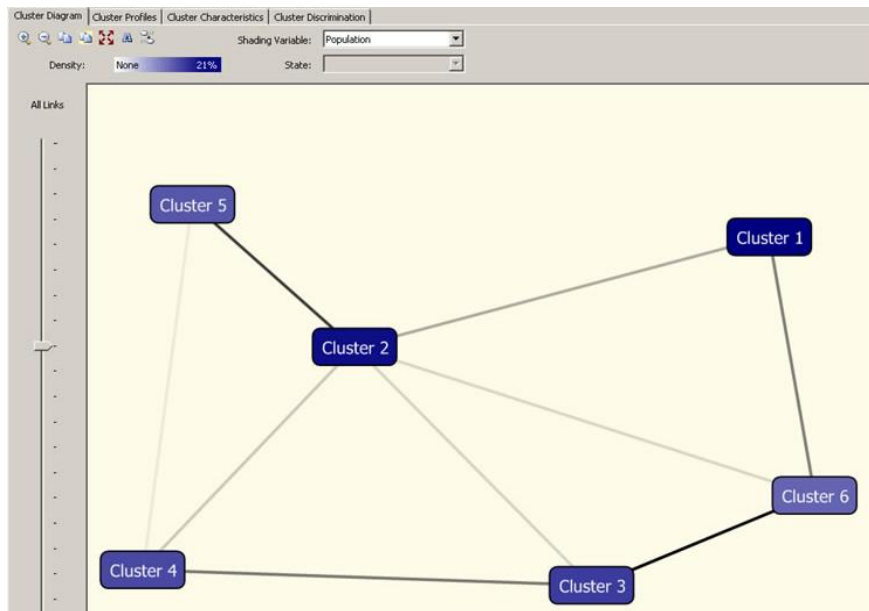


Fig. 4. Store clusters

Discriminating features (attribute/value) pairs can be determined by using the Discrimination browser, as shown in Figure 5.

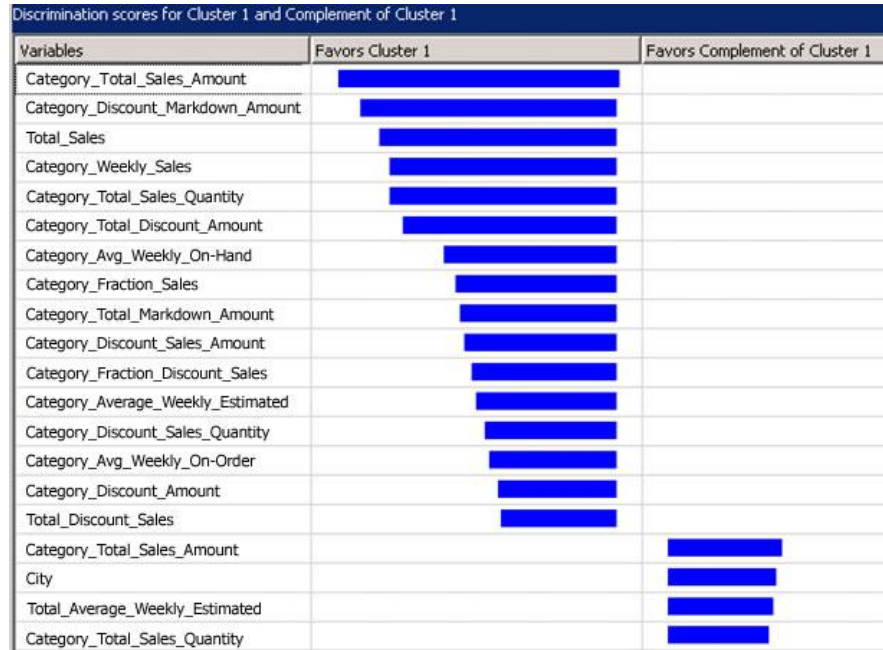


Fig. 5. Cluster model discrimination browser

During the evaluation phase, not only we use tools to evaluate the model accuracy but we also need to discuss the meaning of discovered patterns with business analysts and domain experts.

Sometimes the mining model doesn't contain useful patterns. This may occur for a couple of reasons. One is that the data is completely random. The second reason is that the set of variables in the model is not the best one to use. In this case, we may need to repeat the data-cleaning and transformation step in order to derive more meaningful variables. Data mining is a cyclic process and it usually takes a few iterations to find the right model.

**Phase II: Inventory predictive modeling.** After the store-cluster model that group together stores having similar category sales patterns have been constructed, the next step is predicting whether or not a given spare part will be out of stock one week into the future and two weeks into the future. Prior to building the mining models to make the inventory predictions, we construct modeling datasets for each product of concern.

The dataset used for the inventory predictive model task takes into account weekly sales data for a given spare part across all stores in the supply chain. The *sliding window* strategy to create the dataset for predictive modeling was used. The sliding window strategy typically is a good data preparation strategy when the data has a temporal nature (for example, when predictions are made into the future) and the type of the predictable quantity is discrete (such as Boolean out-of-stock indicators). If there is

sufficient temporal data and the predictable quantity is inherently numeric, time-series modeling may be a preferred strategy.

Typically, there are very few out-of-stock events that occur for a single store and single product. To obtain accurate predictive models, the training data needs to include a sufficient number of out-of-stock events and in-stock events to identify trends differentiating the two. The following data preparation strategy was aimed at achieving a sufficient number of out-of-stock events and in-stock events by considering a given product  $p$  over the entire chain of stores. We included the store cluster label (derived from the store-clustering model) to allow the predictive modeling algorithms to identify trends in out-of-stock behavior that might be different between different store clusters.

For each stores in the retail chain a unique key (store/week identifier) is generated. Some of the attributes which describe the entity are: *current\_week\_on\_hand*, *one\_weeks\_back\_on\_hand*, *one\_week\_back\_sales*, *current\_week\_sales*, *cluster\_label* (from the store-clustering model), *four\_weeks\_back\_sales*, *five\_weeks\_back\_on\_hand*, *two\_weeks\_back\_sales*, *first\_week\_sales\_change*, *one\_week\_oos\_boolean*, *two\_week\_oos\_boolean*, etc.

The data mining algorithms will attempt to identify the pertinent correlations for making accurate predictions. Since the pertinent correlations are not known, we have included all possible attributes in the training dataset. Attributes first/second/third week sales change help to approximate the change in sales week over week. Typically, these types of attributes can be very useful in improving a model's predictive accuracy.

To more objectively evaluate the predictive accuracy of the models, it is common practice to hold out a subset of data and call this the *testing set*. The remainder of the dataset is called the *training dataset*. The data mining models are constructed using the training dataset. Predictions from the model are then compared with the actual values over the testing set.

First a data source is created that specify the database server instance that stores the training and test tables for the spare parts under consideration.

After the data source view is added, a new mining structure is created for the inventory predictive modeling process.

Decision Trees and Neural Network models are built to determine which algorithm produces the most accurate models (as measured by comparing predictions with actual values over the testing set). After an initial mining structure and mining model is built (specifying the input and predictable attributes), other mining models can be added.

In Figure 6, the segment of the mining structure and mining algorithms are shown. Input indicates that the attribute value will be used as an input into the predictive model. PredictOnly indicates that these values should be predicted by the data mining model. Key indicates the column that uniquely identifies the case of interest

| Structure                | InventoryPredictDT       | InventoryPredictNN       |
|--------------------------|--------------------------|--------------------------|
|                          | Microsoft_Decision_Trees | Microsoft_Neural_Network |
| Category_Parts_Cluster   | Input                    | Input                    |
| Current_Week_On_hand     | Input                    | Input                    |
| Current_Week_On_Order    | Input                    | Input                    |
| Current_Week_Sales       | Input                    | Input                    |
| Fifth_Week_Sales_Change  | Input                    | Input                    |
| First_Week_Sales_Change  | Input                    | Input                    |
| Five_Weeks_Back_On_Hand  | Input                    | Input                    |
| Five_Weeks_Back_On_Order | Input                    | Input                    |
| Five_Weeks_Back_Sales    | Input                    | Input                    |
| Four_Weeks_Back_On_Hand  | Input                    | Input                    |
| Four_Weeks_Back_On_Order | Input                    | Input                    |
| Four_Weeks_Back_Sales    | Input                    | Input                    |
| Fourth_Week_Sales_Change | Input                    | Input                    |
| One_Week_Back_On_Hand    | Input                    | Input                    |
| One_Week_Back_On_Order   | Input                    | Input                    |
| One_Week_Back_Sales      | Input                    | Input                    |
| One_Week_OOS             | PredictOnly              | PredictOnly              |
| Two_Week_OOS             | PredictOnly              | PredictOnly              |
| Second_Week_Sales_Change | Input                    | Input                    |
| Third_Week_Sales_Change  | Input                    | Input                    |
| Store_Week_ID            | Key                      | Key                      |

Fig. 6. Out-of-stock mining structure with mining models

#### 4.4. Predictive Modeling Results

The predictive accuracy of mining models were evaluated by examining them over the testing set. There are a few popular tools to evaluate the quality of a model. The most well-known one is the lift chart. It uses a trained model to predict the values of the testing dataset. Based on the predicted value and probability, it graphically displays the model in a chart. The lift chart compares the predictive performance of the mining model with an ideal model and a random model. Figure 7 shows the lift chart for Boolean two-week out-of-stock predictions for the front bulb spare part. The task is to predict a true/false value as to whether the part will be in stock or out of stock two weeks into the future at any store in the chain. The overall predictive accuracy of this model is close to the ideal model.

Table 1 summarizes the predictive accuracies for the four products that were considered in this task. On average, the data mining models can predict whether or not a product will be out of stock one week into the future with 98.68% accuracy. Predictions on whether or not the product will be out of stock two weeks into the future are, on average, 92.46% accurate.

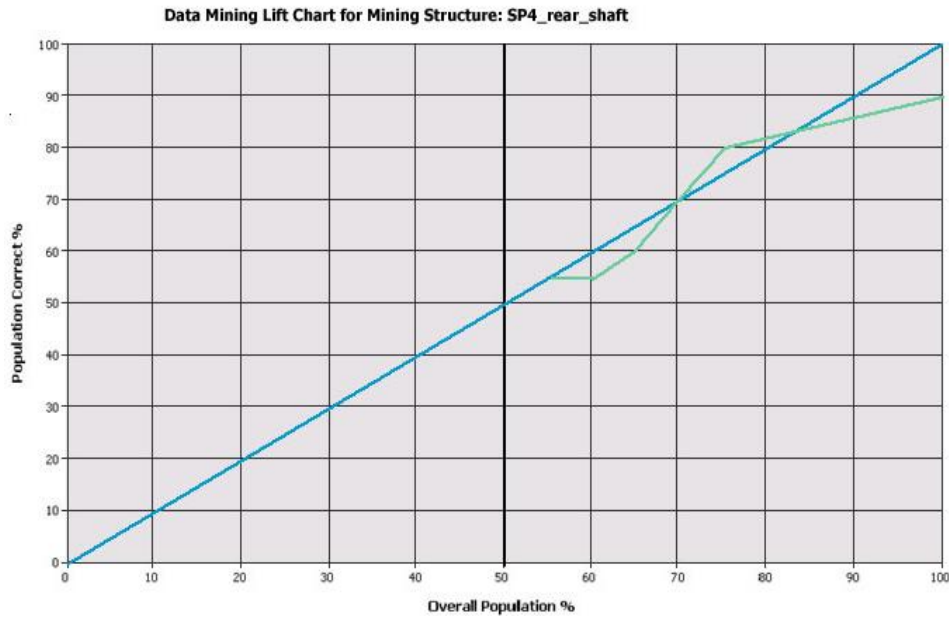


Fig.7. Lift chart for two-week out-of-stock predictions

Table 1.Out-of-stock predictive accuracies for four spare parts

| <i>Product</i>      | <i>Out-of-Stock Prediction</i> |               |
|---------------------|--------------------------------|---------------|
|                     | Week 1                         | Week 2        |
| Product 1           | 98.26%                         | 93.31%        |
| Product 2           | 99.10%                         | 94.12%        |
| Product 3           | 97.65%                         | 89.48%        |
| Product 4           | 99.70%                         | 92.93%        |
| <b>AVG Accuracy</b> | <b>98.68%</b>                  | <b>92.46%</b> |

The presented results show an excellent predictive performance, which makes these BI semantic models very helpful in making inventory-related decisions. The data warehouse scheme (with dimension attributes and measures) and the structure of the data mining models, make this inventory prediction model suitable for various domains (products and industries). Additional, it can be customized or extended with additional dimensions, measures and data mining models to better fit specific problems.

**Sales opportunity.** By using the developed data mining predictive models we can analyze sales opportunities. The formula for calculating the lost sales opportunity for each spare part, was computed by multiplying the number of out-of-stock total store weeks by the two-week Boolean predicted value. Multiplying the out-of-stock predicted values by the percentage of actual sales for the year by the respective retail sale price generates the total sales opportunity. Sales opportunity formula:

$$\text{Yearly increase in sales} = (\# \text{ of total OOS weeks for all stores}) \times (2\text{-week Boolean predicted accuracy}) \times (\% \text{ of actual sales across all stores}) \times (\text{retail price}) \quad (1)$$

Additionally, it is possible to generate profit charts which use input parameters such as: population, fixed cost, individual cost and revenue per individual.

ETL packages can be used to automate the process of making weekly or monthly out-of-stock predictions. These predictions provide the retailer with updated reports on store/product combinations that may be likely to experience an out-of-stock situation.

It is desirable automating not only the process of obtaining out-of-stock predictions on a regular basis, but also that of automating the process of evaluating the performance of the predictive models. The latter task can then be used to determine if the predictive accuracy of the trained data mining models has fallen below an acceptable level. If the trained out-of-stock predictive models fall below a prescribed predictive accuracy, it is likely that the trends and patterns extracted by the data mining models have changed. In this case, new models will need to be constructed and fine-tuned. The process of producing the product/store combinations that may likely experience an out-of-stock situation can be done by implementing and scheduling the following workflow.

## 5. Collaborative BI Portal and Self-Service BI

Traditionally, BI systems were designed and implemented separately of the collaboration systems. However, capability to deliver analytical information to the end-user via standard web technologies, as well as enabling decision-makers to access these information in a unified way, become a critical factor for the success of business intelligence and data mining initiatives. BI portal serve as a virtual desktop providing transparent access to the information objects (models, reports, cubes, spreadsheets, etc.) as described in [30].

In order to provide better user experience and pervasive BI solution, the specialized business intelligence web portal based on the SharePoint platform is designed. It is the integrated, web-based analytical processing solution that enables users throughout the entire supply chain to consume, create and share reports, charts and scorecards, based on OLAP services, cloud data sources, relational databases or web feeds.

Self-service BI is enabled through the incorporation of particular analytical models that enable end-users to create various reports and queries without additional intervention of the IT staff.

BI applications often require specialized propriety client tools and the process of maintenance and modification is time-consuming and difficult. The designed BI web portal offers the standard user interface to easily create centralized place for business analytics. The portal is modular (made of many web parts and services) and enables composition of several data views (modules) in different formats. The main analytical modules are the BI Tree web part which organizes content using a tree structure, the BI Viewer web part for creating views on data, BI Performance Measurement web part, and the BI Data Analysis web part to further analyze or manage the data displayed in a view. Figure 6 shows BI portal with two data views that present two reports for presenting data mining results. The reports are stored in a separate report server and integrated in the portal using the standard web service technologies.



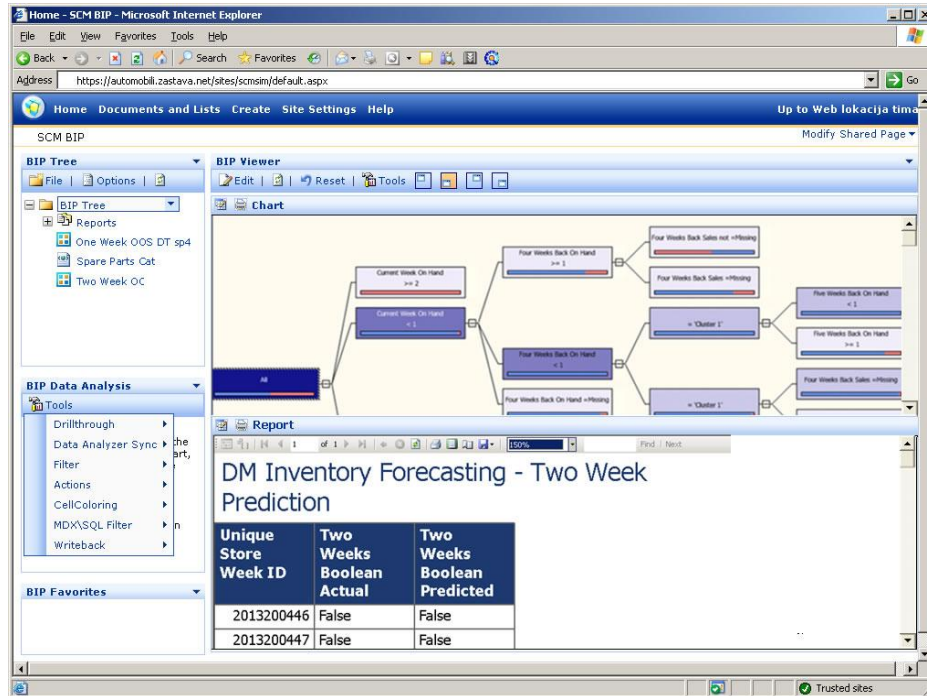


Fig. 8. Business Intelligence Web Portal

The portal can be saved as a template and implemented with out-of-the-box functionality in many locations. The security is enforced through the SSL (Secure Socket Layer) encryption together with the authentication and authorization. User roles (reader, contributor, and administrator) are also supported. These security mechanisms are especially important in the context of the supply chain where many different companies cooperate. This fine-grained authorization mechanism can support different business scenarios, controlling access to the portal, web page, module (web part), and also the single data item.

By implementing the presented BI portal it is possible to deliver data mining results to the right person, any time, via any browser and in a secure manner. Personalization and filtering capabilities enable end users to access information relevant to them. All this features allow supply chain partners to bring more informed decision collaboratively.

## 6. Conclusion

Supply chain as very complex business systems, generate a huge amount of data that can be challenging to integrate, process and analyze. The goals of modern SCM are to reduce uncertainty and risks in the supply chain, thereby positively affecting inventory

control, planning, replenishment and customer service. All these benefits contribute to increased profitability and competitiveness.

In practice, organizations face many challenges regarding the inventory management that include uncertainty, data isolation, local decision making, and problems with data sharing. Most of the existing analytical systems fail to provide the required level of flexibility, performance, and collaboration.

In this paper, the collaborative business intelligence semantic model is presented. The main advantages include better flexibility, scalability, integration, performance, usefulness, and collaboration capabilities.

Based on this BI semantic model, the predictive inventory management analytical solution was designed. It integrates and consolidates all inventory relevant data and uses business intelligence technologies like data warehousing and data mining, to perform accurate forecasts and finally to deliver derived knowledge to the business users via web portal. An approach with data warehouse enables data extraction from different sources and design of integrated data storage optimized for analytical tasks such as data mining.

The presented out-of-stock predictive model is tested with real automotive data set and demonstrated excellent accuracy for one week and two week forecasting. This information can be very useful when making inventory planning and replenishment decisions which can ultimately result in more sales, decreased costs and improved customer service level.

Finally, delivering this information to the right users, in the right format, and within the collaborative environment enables proactive and coordinated decision making.

**Acknowledgement.** Research presented in this paper was supported by Ministry of Science and Technological Development of Republic of Serbia, Grant III-44010, Title: Intelligent Systems for Software Product Development and Business Support based on Models.

## References

1. Turban, E., King, D., Lee, J., Liang, T-P., Turban, D.: *Electronic Commerce 2012: A Managerial and Social Networks Perspective*. Pearson Education, London, UK. (2012)
2. Stefanovic, N., Stefanovic, D.: *Supply Chain Business Intelligence – Technologies, Issues and Trends*. In: Bramer, M. (ed.): *Artificial Intelligence: An International Perspective*. IFIP State of the Art Series, Lecture Notes in Computer Science. Springer-Verlag, 217-245. (2009)
3. Lee, H. L., Padmanabhan, V., Whank, S.: *The Bullwhip effect in supply chains*. *Sloan Management Review*, Vol. 38, 93-102. (1997)
4. Sethi, P. S., Yan, H., Zhang, H.: *Inventory and supply chain management with forecast updates*, Springer Science, USA. (2005)
5. Barry, A. J., Linoff, S. G.: *Data mining techniques*, 2nd edition. Wiley Publishing, USA. (2005)
6. Syntetos, A. A., Boylan, J. E., Disney, S. M.: *Forecasting for inventory planning: a 50-year review*. *Journal of the Operational Research Society*, Vol. 60, 149-160. (2009)
7. *Predictive inventory management: Keeping your supply chain in balance*. IBM (2012). [Online]. Available: <ftp://ftp.boulder.ibm.com/common/ssi/ecm/en/ytw03260usen/YTW03260USEN.PDF>

8. Syntetos, A. A., Nikolopoulos, K., Boylan, J. E.: Judging the judges through accuracy-implication metrics: The case of inventory forecasting. *International Journal of Forecasting*, Vol. 26, No. 1, 134-143. (2010)
9. Chandra, C., Grabis, J.: Application of multi-steps forecasting for restraining the bullwhip effect and improving inventory performance under autoregressive demand. *European Journal of Operational Research*, Vol. 166, No. 2, 337-350. (2005)
10. Liang, W-Y., Huang, C-C.: Agent-based demand forecast in multi-echelon supply chain. *Decision Support Systems*, Vol. 42, No. 1, 390-407. (2006)
11. Boylan, J. E., Syntetos, A. A.: Spare parts management: a review of forecasting research and extensions. *IMA J Management Math*, Vol. 21, No. 3, 227-237. (2010)
12. Hua, Z. S., Zhang, B., Yang, J., Tan, D. S.: A new approach of forecasting intermittent demand for spare parts inventories in the process industries. *Journal of the Operational Research Society*, Vol. 58, 52-61. (2007)
13. Bala, P. K.: Purchase-driven Classification for Improved Forecasting in Spare Parts Inventory Replenishment. *International Journal of Computer Applications*, Vol. 10. No. 9, 40-45. (2010)
14. Dhond, A., Gupta, A., Vadhavkar, V.: Data mining techniques for optimizing inventories for electronic commerce. In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*. Boston, MA, USA, 480-486. (2000)
15. Symeonidis, L. A., Nikolaidou, V., Mitkas, A. P.: Exploiting Data Mining Techniques for Improving the Efficiency of a Supply Chain Management Agent. In *Proceedings of the IEEE/WIC/ACM International conference on Web Intelligence and Intelligent Agent Technology*. Hong Kong, 23-26. (2006)
16. Seifert, M., Eschenbaecher, J.: Predictive performance measurement in virtual organisations. In: Luis, C. M. (ed.): *Emerging Solutions for Future Manufacturing Systems*. Springer, Boston, MA, USA, 299-306. (2005)
17. Derrouiche, R., Holimchayachotikul, P., Leksakul, K.: Predictive performance model in collaborative supply chain using decision tree and clustering technique. In *Proceedings of the 4th International Conference on Logistics (LOGISTIQUA '11)*. IEEE Computer Society Press, Chicago, Ill, USA, 412-417. (2011)
18. Stefanovic, N.: Proactive Supply Chain Performance Management with Predictive Analytics. *The Scientific World Journal*, Vol. 2014, 1-17. (2014)  
doi:10.1155/2014/528917
19. Aruldoss, M., Travis, L. M., Venkatesan, V. P.: A survey on recent research in business intelligence. *Journal of Enterprise Information Management*, Vol. 27 No. 6, 831 - 866. (2014)
20. Sanders, N. R.: *Big Data Driven Supply Chain Management - A Framework for Implementing Analytics and Turning Information into Intelligence*. Pearson Education, Inc. Upper Saddle River, New Jersey, USA. (2014)
21. Waller, M. A., Fawcett, S. E.: Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management. *Journal of Business Logistics*, Vol. 34, 77-84. (2013)
22. Trujillo, J., Maté, A.: Business Intelligence 2.0: A General Overview. In: Aufaure., M-A., Zimányi, E. (eds.): *Business Intelligence*. Springer Berlin Heidelberg, 98-116. (2012)
23. Chen, H., Chiang, R. H. L., Storey, V. C.: Business Intelligence and Analytics: From Big Data To Big Impact. *MIS Quarterly*, Vol 36. No. 4, 1165-1188. (2012)
24. Berthold, H., Rosch, P., Zoller, S., Wortmann, F., Carenini, A., Campbell, S., Bisson, P., Strohmaier, F.: An architecture for ad-hoc and collaborative business intelligence. In *Proceedings of the 2010 EDBT/ICDT Workshops (EDBT '10)*. ACM, New York, NY, USA, 1-6. (2010)
25. Rizzi, S.: Collaborative Business Intelligence. In: Aufaure., M-A., Zimányi, E. (eds.): *Business Intelligence*. Springer Berlin Heidelberg, 98-116. (2012)

26. LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., Kruschwitz, N.: Big Data, Analytics and the Path From Insights to Value. MIT Sloan Management Review, Boston (2011). [Online]. Available: <http://sloanreview.mit.edu/article/big-data-analytics-and-the-path-from-insights-to-value>
27. Stefanovic, N., Stefanovic, D.: Integrated and interactive software solution for knowledge-based supply network design. Computer Systems Science & Engineering, Vol. 28. No. 1, 5-23. (2013)
28. Rad. R.: SQL Server 2014 Business Intelligence Development. Packt Publishing, Birmingham, UK. (2014)
29. Stefanovic, N., Stefanovic, D., Radenkovic, B.: Application of Data Mining for Supply Chain Inventory Forecasting. In: Ellis, R., Allen, T., Petridis, M. (eds.): Applications and Innovations in Intelligent Systems XV, Springer London, 175-188. (2008)
30. Stefanovic, N., Stefanovic, D.: Supply Chain Performance Measurement System Based on Scorecards and Web Portals. Computer Science and Information Systems, Vol. 8, No. 1, 167-192. (2011).

**Nenad Stefanovic** completed his BSc in Mechanical Engineering with the University of Kragujevac, where he also got his MSc in Information Systems. He received his PhD in Electronic Business from the Faculty of Organizational Sciences, University of Belgrade. Nenad has a strong background in industry sector, where he worked as Chief Software Architect, Chief System Engineer and Project Manager (ICT Department, Zastava automobiles), and also as ICT Operations and Architecture Manager (ICT Department, FIAT Auto Serbia). He works as a professor at the Faculty of Science (Department of Mathematics and Informatics) and the Faculty of Technical Sciences (IT Department), University of Kragujevac. Nenad Stefanovic is a member of several national and international ICT and business related societies and research projects. He is the vice-chair of the IFIP TC12 (Technical Committee 12 - Artificial Intelligence) and national representative in the IFIP TC2 (Technical Committee 12 – Software and Theory). He was the President of the JUSK (Serbian National Organization for Quality) Software Committee for several years. Nenad holds the highest category for researchers (T1), assigned by the Serbian Ministry of Science. Nenad has published more than 70 scientific papers and book chapters in national and international journals, conference proceedings and books. His research interests include business intelligence, supply chain management, electronic business, information systems, web technologies, project management and simulation.

*Received: November 1, 2014; Accepted: February 15, 2015.*