10-2019

Near Real Time Business Intelligence Framework using R Shiny

Bruno Fernandes

Mateus Mendes

Jorge Alexandre Almeida

Near Real Time Business Intelligence Framework using R Shiny

Bruno Fernandes, Polytechnic Institute of Coimbra – ISEC, Portugal, a21270987@alunos.isec.pt

Mateus Mendes, Polytechnic Institute of Coimbra – ESTGOH and Institute of Systems and Robotics of the University of Coimbra, Portugal, mmendes@isr.uc.pt

Jorge Alexandre Almeida, Polytechnic Institute of Coimbra – ISEC, Portugal, jalmeida@isec.pt

Abstract

Modern industry deals with large amounts of data, which is often difficult for humans to process and use for decision making. Industry 4.0 proposes the automation of different procedures in enterprises, aiming to reduce human errors, operation time and costs. That includes analysis of different operation parameters in near real time, in order to facilitate management to make the right decisions at the right time. That requires the use of tools that are simple and fast to use and provide the necessary information. The present paper describes an architecture of a Business Intelligence system proposed for a Telecommunications software company. The system draws information from a proprietary ERP and is all developed using free open source software. The architecture proposed uses the power of R for statistical computing, data mining and artificial intelligence. Financial information is shown in a dashboard in near real time.

Keywords: Business Intelligence; Financial Information Dashboard; Data Mining; Neural Networks

1. INTRODUCTION

Nowadays, companies are faced with enormous amounts of information, produced internally and externally at a fast pace. That information is derived from their internal processes, such as production, accounting, logistics, commercial, and financial area, as well as their external relationship with customers, suppliers and banking entities. Storing and analyzing data derived from those relationships, companies know that they have information vital to their survival in the market, that can bring them competitive advantages. Most of that information is supported by enterprise management systems, called ERP (Enterprise Resource Planning). The ERP's main task is to consolidate company information into a single database, combining the functions of production control, projects, inventory, purchases, human resources, billing and treasury, among others, and thus eliminating the difficulty of dealing with data stored in various systems.

The amount of information stored over time can be very large, even for small enterprises. Many companies still process the data using spreadsheets, but that quickly becomes impractical for big companies or old companies which want to process data stored over several years. Therefore, the results obtained are only part of what could be discovered using more powerful tools. To overcome the difficulty and extract a good amount of knowledge about past events, there are more powerful technologies and tools that form the basis of Business Intelligence (BI) systems. Those include tools of Data Warehouse (DW), On-Line Analytical Processing (OLAP), and Data Mining (DM) (Shim et al., 2002).

1.1. Data Warehouse

In the last few years, databases have been improved for operational and analytical data handling, with users tending to the operational level, even though they also need analytical tools. That is different from operational databases. DWs are designed to store gigabytes and even terabytes of information and can sometimes be divided into smaller logical drives called Data Marts (DM). According to Kimball & Ross (2013), DM are DW but contain data from a particular area or department of the organization. Data warehouses are considered as key tools of the BI infrastructure. To be effectively implemented, they need to be embedded in a complete BI architecture. Data warehouses have a solution for data access, reconciliation problems, and quality. To achieve this goal and power a data warehouse, significant work must be done by a process called ETL (Extract-Transform-Load) commonly called (Linden, 2019).

1.2. OLAP

According to Zeng, Xu, Shi, Wang & Wu (2006), OLAP refers to the techniques of execution and complex analysis of information, stored in a DW or not. Those techniques transform raw data into information, which is more easily perceived, analyzing the various dimensions and combinations, aiming to recognize trends in the data. Costa & Santos (2012) affirm that OLAP systems are used to analyze DW. They allow the creation of cubes analyzing the information generated from several angles. After the creation of cubes, those cubes represent integrated data with some measure of interest, so they can have several dimensions to analyze and represent a particular business.

1.3. Data Mining

Data mining is the action of extracting information from a database, using sophisticated analysis techniques such as those coming from Artificial Intelligence, Statistics and Mathematics, in order to produce valuable knowledge used in the strategic management of an organization, according to Bispo & Cazarini (1998) and Turban et al. (2010). Applying the most appropriate Data Mining techniques to achieve the goals defined is considered an information modelling phase. The main techniques used are decision trees, artificial neural networks and rules of induction. The parameters of each model are adjusted in order to optimize the results (Bispo & Cazarini, 1998).

Neuronal networks are modified regression models that make approximations, according to Abad & Chaparro (2016). They are networks trained using inductive reasoning, with a set of input and output data. They consist of nodes and connections, where each connection has a weight which is adjusted during the training phase. They have the advantage of being able to abstract an interpretation of complex data in order to represent patterns or trends. The rules of induction are an exploratory method, with the purpose of finding models or categories of data that are grouped in a natural way.

1.4. Structure of the paper

Section 2 describes the related work. Section 3 describes BI model proposed. Section 4 shows results produced by the system. Sections 5 and 6 draw some conclusions and perspectives of future work.

2. RELATED WORK

In a given period of time, a company generates a large amount of data related to its production, billing, stocks, employees, materials, products, etc. The raw data, will have little utility to withdraw any conclusion or extract any knowledge.

According to Mussa, Souza, Freire, Cordeiro and Hora (2018), a BI tool has as main objective to provide information in the right way and in the right time so that the company can apply the best decisions in a faster and more reliable way. It is a set of tools and applications that allow decision makers to organize, analyze, distribute and thus act on information relevant to the company's business.

BI systems at a strategic level enable to accurately define objectives and monitor the achievement of such objectives, allowing to make comparative reports on an organization's profitability or the effectiveness of the distribution channels by simulating or predicting future results. At a tactical level they offer a basis for optimizing marketing, sales and financial decisions in order to help achieve strategic objectives in a more effective way. At the operational level, BI systems are used to perform *ad-hoc* analysis and answer questions related to the current operations of the organization's activity, typically focusing on suppliers and customers (Olszak et al. 2007).

Nowadays, BI systems are being regarded as a powerful set of tools which allow to create value for a company. In that context, the present paper aims to briefly summarize some architectures proposed by different authors for different activities, with diverse goals and using various tools.

According to Zeng et al. (2006), more and more business sectors are deploying advanced BI solutions to increase their competitiveness, due to their importance for the efficient functioning of the organizations in managing their resources and developing the businesses. Selecting and implementing a BI tool can be a time-consuming and difficult task. There are now producers that offer a wide range of solutions, from low-tech products that offer just simple reporting, to sophisticated platforms. Some platforms are open source, others are proprietary software.

According to Howson, Richardson, Sallam and Kronz (2019), modern (BI) platforms are characterized by easy-to-use tools for flowing from a complete analytical workflow, from data preparation to visual appearance. They stand out from traditional BI platforms, not requiring significant involvement of IT teams to predefine data storage models.

A comparison of the classification of BI platforms is presented in Figure 1. The figure plots the platform’s ability to execute as a function of the completeness of vision. Platforms like Tableau, Microsoft and other emerging vendors offer a restricted set of features, but are widely used for a variety of BI functions and custom reports. In general, they are easy to use and deploy.

The market leader, Microsoft, offers data preparation, visual-based data discovery, interactive dashboards and enhanced analysis through a single product, Power BI. Power BI Desktop can be used as a standalone, free self-service analysis tool. Power BI can be used by people with minimal skills in data science, among many

others. Important script items include support for existing Reporting Services, a common data model, and open data preparation with data streams.

BI models are present in all sectors of activity. This article intends not only to deal with models of BI implementations in software companies, but also to identify some BI implementations in the most diverse sectors.

A BI system is never implemented by a single technology, product, or vendor. There are several models and implementations of BI systems. Costa and Santos (2012) propose a model that allows an organization to obtain information that supports decision making at all levels of planning, whether strategic, tactical or operational. By leveraging the entire information history of its database, the BI solution allows the organization to access information on purchases and sales of products, accounting and financial information, as well as current accounts of suppliers. Using data from the ERP, it stores the necessary information in a local database management system. The author-defined architecture consists of a first server with its ERP database, a second server with the Data Warehouse, the ETL process and the OLAP and reporting servers. Report analysis is supported by a web environment and a spreadsheet software. Because of the complexity of the database, the authors chose to create views of the tables considered essential so as not to overload the BI system. The Data Warehouse environment, besides allowing the multidimensional modelling of the data, may allow the introduction of Data Mining techniques. The authors implemented three interface scenarios, i) using a front end reporting solution; ii) using a spreadsheet software; and iii) using the Business Intelligence tool Pentaho.

Figure 1 - Magic Quadrant for Analytics and Business Intelligence Platforms according to Howson et. al (2019)

The authors, Christian, Akujobi, Saha and Korzaan (2018) report the implementation of a business intelligence and analytics system in a software production company for the gaming industry using Tableau. The first step was to clean and migrate the data from the different databases, using R programming for this integration. The software was used in this case to aid in the analysis of statistics and repeater data of complex algorithm patterns. The second step consisted in designing the data model and defining its visualization using Tableau.

Another model of BI implementation is proposed by Ali, Nassif and Capretz (2013), more specifically in the area of healthcare. With a significant increase in infection rates, entailing very high costs to the organization, the intention of Ali *et al.* was to define and recommend an appropriate data and reporting management structure to support the decision-making strategy. The authors argue that implementing a robust framework to obtain, manage and report data and information is a key success factor in those projects, with the aim of reducing the number of infections and the associated costs. The BI implementation project uses the OLTP (On-Line Transaction Processing) tools responsible for handling routine data that is generated daily through the company's information systems. It also uses analytics tools for decision support, allowing to solve problems in a simple and economical way through a straightforward and integrated view of the data. The authors defined the following development steps: 1) Design the database; 2) Define the ETL process; 3) With the desired data, the structure of the OLAP cubes was drawn, creating and displaying the data source, associated dimensions, the multidimensional cube and the KPIs (Key Performance Indicators) to display monthly metrics; 4) Define and construct information reports; and 5) Add Data Mining components. With healthcare organizations experiencing difficulties in extracting valuable information from their data, the tool allows them to increase strategic thinking through modifications to the classic information tools for BI solutions.

Artificial neural networks have already been used data mining processes and BI solutions, and numerous examples can be found in the literature. Nwuf *et al.* (2017) defined a neural network model to analyze the performance of a combustion engine. The model was designed to predict fuel consumption, maximum pressure achieved and engine efficiency by using the engine speed register, fuel ratio and exhaust gas temperature as input. They found that the neural network model provided accurate information for the analysis of complex equipment and was considered useful for predicting the performance of combustion engines. Rodrigues *et al.* (2019) use artificial neural networks to model the condition of diesel engines' oil. Based on twenty one variables that represent the oil condition, the authors conclude that the ANN can learn and make decisions similar to human experts and, therefore, the process can be automated.

3. DESCRIPTION OF THE BI MODEL PROPOSED

The present paper describes a BI solution implemented in a Portuguese company that operates in the field of software development for the mobile telecommunications and mobile internet sectors. The company has verified over the years a remarkable growth in revenue, number of employees and facilities. It is classified as a big company, according to the classes defined in the Portuguese Decree Law no. 158/2009 of July 13, which approves the Portuguese accounting standardization system. It has continually firmed new partnerships and penetrated new markets, namely with telecommunications companies with a strong presence in the national

and European markets. The focus of the company will be to maintain this trend of evolution, conquering emerging markets in the telecommunications sector. Due to its continually growing size, and the projected growth, a support system at the decision-making level, through the implementation of a BI system, is appropriate and necessary. In order for managers to simplify the processes when making strategic management, the company aims to suppress the spreadsheets used to analyse cost and performance indicators, with sometimes extensive information. To address these constraints, the company needs a system that gives historical and current data information available in the ERP database, and which also makes projections for the future and thus assists the management in decision making.

The development of the BI solution proposed started with meetings with the management of the company, as well as the technical department, in order to understand the requirements of the BI model aimed by the management and the technical infrastructure already in place.

The company currently uses an ERP system as an integrated management system, including all accounting, financial and human resources information. The management showed interest in a BI solution that would show, with a time gap as small as possible, the evolution and trends of revenue and expenses, per project, company department and in total.

The goal is therefore to extract the necessary billing information, market positioning, expenses with employees and total expenses, from the ERP, to use it for the BI system for further analysis and calculation of the economic and financial indicators. The architecture of the solution includes open source software that allows to extract and transform the information, being prepared to receive artificial intelligence algorithms, namely neural networks, for making the projections of future trends. Sang, Xu and Vrieze (2016), and the Free Software Foundation[1], define Open Source Software (OSS) as the software that can be freely used, modified and distributed by anyone. Therefore, OSS can always be improved, so it is the preferred choice for the present system.

The presentation of the information allows the visualization of the data in real time. It is generated through the same software where the user interface is created as dashboards, containing tables, graphs and iterative tools of multidimensional analysis that allow to evaluate the current state of the business. The BI tool was expected to provide treatment and analysis of the information considered relevant for those responsible for top management, in order to help them make a more efficient and effective decision making.

[1] http://www.fsf.org (last visited 2019-05-09).

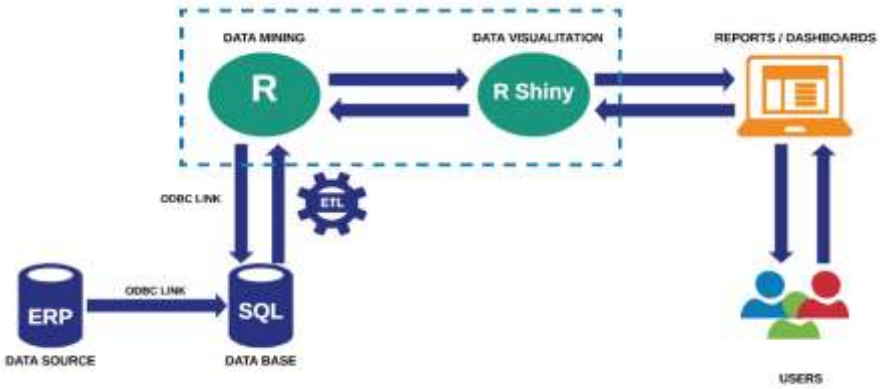

Figure 2 - Arquitecture of the Business Intelligence system developed

After the initial phase in defining the company's requirements, the decision was to make a copy of the relevant information from the ERP database to a separate database, where data could be processed using Rstudio. RStudio software is an open source tool, which meets all the requirements intended for the system. R is a programming language and computational environment, specialized in data manipulation, statistical analysis and graphical visualization. It is available for different operating systems. For user interface, RStudio software was chosen. It has a package name RShiny, which facilitates the creation of interactive web applications that can be used to display data in an interactive way with innovative visualization features. Figure 2 shows the architecture of the system, highlighting the components and technologies used for its construction.

The above mentioned architecture comprises a server that stores information copied from the ERP. Copies of the data from the ERP to the server are done automatically every five minutes, through an ODBC link. This method ensures that the ERP information is available in near real-time for the BI system, while the BI data and the ERP data are separated for integrity reasons. The BI system has read-only access to the ERP database, while the ERP has no access to the BI database. RStudio has then access to the BI database through another ODBC connection, which grants read and write access, as shown in Figure 2.

After the dataset is extracted from the database, it is processed in R and sent to R Shiny. R Shiny prepares the processed data for visualization in the dashboard. The dashboard offers different visualization options to the user, so communications between R, R Shiny and the user are bi-directional.

Figure 3 shows the data mining process in more detail. Selected data are extracted from the database and fed to an R module responsible for applying the necessary statistical and artificial intelligence algorithms. This module is where the financial parameters are calculated, based on the information retrieved from the database, and the charts are produced. The module is also prepared to include a feed forward neural network, namely supervised predictive models, available in R. The neural network planned is three-layered model. The number of neurons will be determined experimentally in future work. The network will be trained and used to project the future growth of the company's billing. In the process of handling historical billing data, users select the year and billing grouped by quarter within each year. This way the network is expected to provide a good result

for the forecast of billing for the next quarters. The output of this module is then forwarded to the next module, which prepares the output and user interface. The user interface was designed in R Shiny, as stated previously.

The user interface was designed in the form of a dashboard, accessed through a web browser. A dashboard is a display area that can contain various elements such as tables, graphs, indicators or maps. It is a visual representation that makes it possible to display the performance of the organization, according to the latest available data. Barros (2013) claims the dashboards assume a fundamental importance in the supervision of the organization, supporting the decision making process. They should be organized graphically, and should have graphical views appropriate to the information that is intended to display. In the construction of web applications and according to Konrath et al. (2018), Shiny presents two scripts that are written to a directory of the R environment. The first script, ui.R, refers to the User Interface (UI). The UI has the function of controlling the layout and the appearance of the dashboard. The second script is the server and is named server.R.

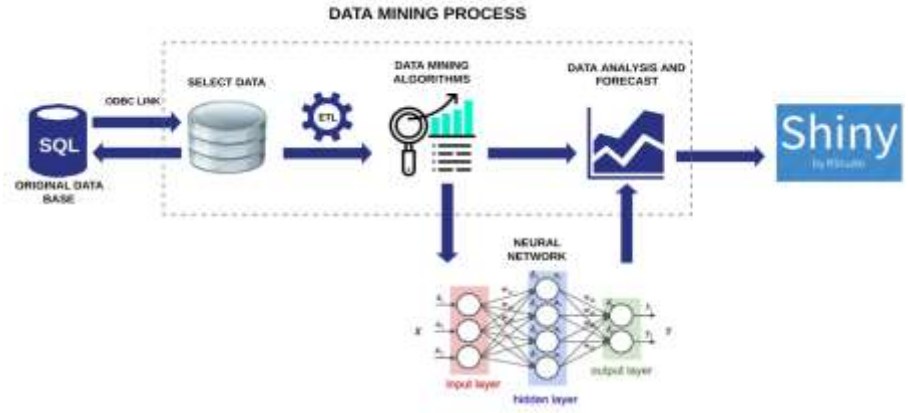

Figure 3 - Details of the data mining process

4. OUTPUTS OF THE BI SYSTEM

Figure 4 shows a screenshot of an example of a dashboard produced by the BI system proposed. One of the objectives of the prototype is to provide a BI tool that makes it possible to analyze data in an easy and intuitive way. As shown in the figure, in the first tab there are charts that display the evolution of billing per year, by market, personnel expenses and revenues over the years. There is a second tab that shows employee distribution, facilitating the analysis of geo-referenced data. The third tab will show forecasts, produced using AI algorithms. The AI algorithms are presently being implemented. They will allow to make projections for the future.

Figure 4 - Example of the dashboard produced by the BI system, showing financial reports

The first tab, called “Financial Reports”, shows on top the situation of the company at a glance, regarding income and expenses. The green boxes show the invoicing and revenue for the current year, compared to the all time highest invoicing and revenue. The yellow boxes show the costs of employees and the total costs for the current year, compared to the previous year. The second row shows the evolution of invoicing and revenue, and their distribution per market.

The second tab shows a map of the distribution of the company’s human resources, as depicted in Figure 5. This visual information allows, if the company decides to change offices, define the new location according to the largest number of employees in a given area, reducing travel expenses. Currently the largest workforce is in Coimbra, with the second largest in Lisbon and a small representation in the north of Portugal.

Figure 5 - Example of the dashboard produced by the BI system, showing employee distribution

5. MAIN CONTRIBUTION AND DISCUSSION

The BI tool described has been proposed to financial and administration users for validation. In the future, it is expected that it will be used by the department of human resources and operations with information applicable to each department. As advantages of this tool, one is the use of open source software, being possible to integrate other programming languages, which offer different software packages. The architecture proposed can be applied to any company, as long as it uses the same ERP or the necessary changes are implemented. The neural models can be trained with existing data, as long as the format is preserved. As disadvantages, it should be noted there is no official support for the R language and R Shiny.

6. CONCLUSION AND FUTURE WORK

A Business Intelligence system is an important asset in organizations, to help managers make the best strategical decisions. Modern systems should provide not only the tools to analyze the past, but also help management understand what will probably happen in the future based on the present and past data.

The present paper describes a BI solution, proposed for a big telecommunications software company. The solution proposed imports the relevant data from the ERP, processes the data using R Studio and shows the results in a dashboard designed in RShiny. It is a simple, powerful and inexpensive architecture. The system for analyzing past information is already implemented, showing the company income and expenses at a glance, as well as the distribution of human resources. The module for making projections for the future, using neural networks, is currently under development.

Author contributions: Bruno Fernandes developed the software and outlined the first draft ot the paper. Jorge Almeida and Mateus Mendes proposed the architecture of the system, supervised the project and contributed equally to the paper.

REFERENCES

Abad, F. M., & Chaparro, A. (2016). Data-mining techniques in detecting factors linked to academic achievement, https://www.researchgate.net/publication/309033106 (21 April 2019).

Ali, O. T., Nassif, A. B., & Capretz, L. F. (2013). Business Intelligence Solutions in Healthcare A Case Study: Transforming OLTP system to BI Solution, https://www.researchgate.net/publication/259759844 (27 April 2019).

Barros, R. M. P. C. (2013). Dashboarding – Projeto e Implementação de Painéis Analíticos. (Dissertação de Mestrado não editada, Mestrado em Engenharia Informática). Universidade do Minho – Portugal.

Bispo, C. A. F., & Cazarini, E. W. (1998). A nova geração de sistemas de apoio à decisão, https://www.researchgate.net/publication/242619256 (30 April 2019).

Costa, M., & Santos, M. Y. (2012). Sistema de Business Intelligence no suporte à Gestão Estratégica - Caso prático no comércio de equipamentos eletrónicos. Actas da 12ª Conferência da Associação Portuguesa de Sistemas de Informação.

Christian, N., Akujobi, I., Saha, M., Korzaan, M. (2018). ACS: Bringing Business Intelligence and Analytics to a Massive Multiplayer Online Gaming Company. Information Systems Education Journal (ISEDJ), 16(2), pp. 11-17.

Decreto Lei n.º 158/2009 de 13 de julho. Diário da República n.º 133/2009, Série I de 2009-07-1

Howson, C., Richardson, J., Sallam, R., Kronz, A. (2019). Magic Quadrants for Analytics and Business Intelligence Platforms, https://www.gartner.com/doc/reprints?id=1-68720FP&ct=190213&st=sb (10 July 2019).

Kimbal, R., & Ross, M. (2013). The Data Warehouse Toolkit: the complete guide to dimensional modeling. (3 ed.), John Wiley and Sons, Inc.

Konrath, A. C., Silva, S. A., Henning, E., Santos, L. M., Mirandas, R. G., & Samohyl, R. W. (2018). Desenvolvimento de Aplicativos Web Com R e Shiny: inovações no ensino de Estatística. Abakós 6 (2) pp. 55-71.

Linden, I. (2019) 30 years Business Intelligence : from Data Analytics to Big Data. 5th International Conference on Decision Support System Technology – ICDSST 2019, University of Madeira, Portugal, pp. 243-251.

Mussa, M. S., Souza, S. C., Freire, E. F. S., Cordeiro, R. G., Hora, H. R. M., (2018). Business Intelligence in Education: An application of Pentaho Software. Revista Produção e Desenvolvimento 4, (3), pp. 29-41.

Nwufo, O. C., Okwu, M. O., Nwaiwu, C. F., Igbokwe, J. O., Nwafor, O. M. I., Anyanwu, E. E. (2017). The Application of Artificial Neural Network in Prediction of the Performance of Spark Ignition Engine Running on Ethanol-Petrol Blends. International Journal of Engineering and Technologies (12), pp 15-31.

Olszak, C. M., & Ziemba, E. (2007). Approach to Building and Implementing Business Intelligence Systems, https://www.researchgate.net/publication/255612097 (21 April 2019).

Rodrigues, J., & Costa, I., & Farinha, J. T., & Mendes, M., & Margalho, L. (2019). Modelling Diesel Engine Oil Condition Using Artificial Neural Networks, eMaintenance 2019, Stockholm, Sweeden, May 2019.

Sang, G. M., Xu, L., & Vrieze, P. (2016). Implementing a Business Intelligence System for small and medium-sized enterprises, https://www.researchgate.net/publication/313037031 (1 May 2019).

Shim, J.P., Warkentin, M., Courtney, J.F., Power, D.J., Sharda, R., & Carlsson, C. (2002). Past, present, and future of decision support technology. Decision Support Systems (931), pp. 1-16.

Turban, E., Sharda, R., & Delen, D. (2010). Decision Support and Business Intelligence Systems (9 ed.), Prentice Hall.

Zeng, L., Xu, L., Shi, Z., Wang, M., & Wu, W. (2006). Techniques, process, and enterprise solutions of business intelligence. IEEE International Conference on Systems, Man, and Cybernetics (6), pp. 4722-4726.