Interactive BI Platform for Business Performance Analysis

Alina Simón Cuevas\textsuperscript{1,2}  
MSc. in Computer Science  
alina.simoncuevas@ucalgary.ca  
(587) 966-4111

Maité Torres Sánchez\textsuperscript{2}  
BSc. in Computer Science  
maitet90@gmail.com  
(786) 678-1398

Lucina García Hernández\textsuperscript{2}  
PhD. in Mathematical Sciences, Computer Science  
lucina@matcom.uh.cu

\textsuperscript{1} Development Data, Development and Alumni Engagement, University of Calgary  
\textsuperscript{2} Faculty of Mathematics and Computer Science, University of Havana

Abstract

Over the last few years, data has become the lifeblood of organizations. Those that harnessed the power of this data by empowering business users found a competitive advantage and were able to innovate faster. These organizations replaced traditional, IT-centric reporting tools with easy-to-use and intuitive modern business intelligence platforms that enable self-service at scale. IT and the business started collaborating to maximize the impact of their data. This approach was developed in a Cuban company, which leads the market in imports, exports, and marketing of products and services, and recognized for software development. Previous Business Intelligence solutions did not support the executives’ independence from IT developers for comprehensive analysis of the data, nor the forecasting of key performance indicators to make tactical and strategic decisions.

This paper outlines the conception, design, and development of a Business Intelligence Platform based on the integration of enterprise analytical scenarios to offer novel and heterogeneous data presentations in addition to data mining models for predicting business performance. Focused on an Enterprise Data Warehouse considering multidimensional and tabular analytical approach, it has evolved as consecutive business intelligence solutions over last years. Microsoft Time Series data mining algorithms were applied for key performance indicators’ forecasting and the results accuracy was evaluated using the SMAPE metric. Functional software and successful experiments validate the global solution.

Keywords — Business Intelligence, Multidimensional Model, Tabular Model, Data Mining, Key Performance Indicator
1. **Introduction**

We live in exciting times from a data analysis point of view. Extracting useful knowledge from data has been the goal of Business Intelligence (BI) since 1958 when Hans Peter Luhn used this term for the first time. Over the last 50 years, the technology has evolved, increasing the manageable amount of data and lowering the related costs. However, it always required professionals who were able to create and refresh the data model, empowering end users with canned reports and data navigation tools (Russo, 2013).

CIMEX Corporation is the largest Cuban company specializing in the import and commercialization of products and services. The primary data to be retrieved for the analysis of user requirements are managed different ERP systems (Enterprise Resource Planning). Existing BI solutions were insufficient for tactical and strategic decision making. One of the main motivations of the authors was to offer an analytic platform to executives which could contribute to more effective and timely use of the information and take advantage of the experiences accumulated over than ten years.

Microsoft provides two alternatives for analytical data models development based on business logic: the multidimensional model and the tabular model. Both models were implemented exploring and evaluating their advantages and disadvantages. Qualitative and experimental comparisons evidenced that the tabular model is more convenient for small data volumes and limited hardware capacities, and on the other hand, multidimensional is more appropriate for modeling more complex solutions and large volumes of data. (Simón, Torres, & García, 2016; Vitt & Cameron, 2012).

One of the main contributions of this research is the first approximation to a data mining model for the key performance indicator's (KPI) forecasting. This proposal was included in the conceptual design of the BI platform and was implemented as part of the computational solution.

2. **Enterprise Business Intelligence Platform**

The solution focuses on the design and implementation of a three-layer Enterprise Data Warehouse (EDW) including data sources (operational data stores), a data repository (relational), and an analytical database. The third layer includes more than fifteen dimensional models of commerce, accounting, finance, and human resources areas. They were developed as cubes in the dimensional models and as
perspectives in the tabular models. Main economic and financial KPIs were designed and implemented taking advantage of the time intelligence functions in the MDX and DAX languages. After evaluating diverse data mining approaches, we consider that time series methods fit conveniently to the prediction of performance indicators. Figure 1 shows the system architecture:

![Enterprise Data Warehouse and BI platform architecture](image)

Fig. 1 Enterprise Data Warehouse and BI platform architecture.

Using Business Intelligence Semantic Model (BISM) allows you to combine data from multiple sources into a common model regardless of how the results are presented. Interactive tools are offered to allow users to explore data with self-service functionalities and provide rich visual presentation and analysis. Users can create reports and dashboards that include charts, widgets, KPI metrics, pivot tables, and tabular view components. A web portal in SharePoint Server 2010 encourages collaboration among executives.

3. **Key Performance Indicators Forecasting**

The *Cross Industry Standard Process for Data Mining* (CRISP-DM) was selected for the knowledge discovery development project (Brown, 2014; Provost & Fawcett, 2013). Cleaned and reconciled data in the relational DW constitutes the data source for the designed mining models. For KPIs forecasting, the Microsoft Time Series regression method was selected considering the goal of predicting future values based on previously observed monthly values, such as income, expenses, profits, sales, purchases, etc. This method implements three algorithms: ARTxp (*AutoRegressive Tree with Cross Prediction*), an
autoregressive tree model for representing periodic time series data developed by Microsoft Research; ARIMA (AutoRegressive, Integrated Moving Average), applying the differentiation approach proposed by Box-Jenkins to convert a series non-stationary in stationary; and Mixed (a combination of both), blending the results to improve prediction accuracy (MacLennan, Tang, & Crivat, 2009). Different parameters combinations were tested on the algorithms to find the most accurate data mining model.

The evaluation of the obtained results for each algorithm is an essential phase during the forecasting. SMAPE (Symmetric Mean Absolute Percent Error) was chosen as the evaluation metric, one of the most recognized metrics to evaluate the predictions in the areas of economy and finances (Montgomery, Jennings, & Kulahci, 2015; Tofallis, 2015).

4. Results

The platform was validated through several experiments using six years historical data from more than 50 subsidiary companies. Three of the most relevant experiments are detailed below and were developed using four production servers in the holding company infrastructure.

The first experiment verified the process of populating the relational DW measuring the execution time and database size. More than 70 million records were processed in 5 hours occupying 6.5 Gb. A monthly average of 4.5 minutes is estimated to monthly process 987 thousand records growing the database in 93 Mb each month. These execution times are considered acceptable considering the volume of data handled and the necessary processing to maintain the referential integrity and scenarios integration. Regarding the analytical database, the most interesting result was the processing time of the tabular model, which took only 11 minutes and used 1.1 Gb RAM memory. This fact responds to the xVelocity in-memory analytics engine which employs techniques such as columnar storage, state-of-the-art compression, in-memory caching and highly parallel data scanning and aggregation algorithms (Ferrari, 2012).

The second experiment successfully tested the analytical platform scope regarding self-service, dynamism, and easiness to obtain timely and accurate information. The visual enrichment was based on the diversity of options for business users. Different queries were run using Power View for SharePoint, Power Pivot for Excel y Power BI.
The third experiment demonstrated the validity of the instrumented mining process using the SMAPE metric on income forecasting results. The Time Series algorithms ARTxp, ARIMA, and Mixed (ARIMA 50%, ARTxp 50%) were applied selecting twelve months historical data from January 2015. The results reveal that ARTxp algorithm provides the best forecast for the income indicator experimented with 3, 6, 9 and 12 months of historical values.

5. CONCLUSION

The present research approached the creation of an interactive platform for the analysis and forecasting of key performance indicators for a large trading company according to the Business Intelligence paradigm. The proposal addressed the integration of the business scenarios into an Enterprise Data Warehouse based on multidimensional and tabular models. The usage of data mining techniques for prediction of financial indicators contributed to better preparation in the present and decreased future risks. Business users gained the ability to perform analytics and drive decisions in a self-service and collaborative environment stimulating the knowledge management in the company.

REFERENCES


