Analyzing sales to generate strategies with IBM Cognos Analytics

Analizando ventas para generar estrategias con IBM Cognos Analytics

CASTORENA-PEÑA, Jesús Abraham†*, SILVA-AVILA, Alicia Elena, DOMÍNGUEZ-LUGO, Alma Jovita and MARTÍNEZ-CASTRO, Jehú Efraín

Universidad Autónoma de Coahuila. Barranquilla s/n, Col. Guadalupe C.P.25750

ID 1st Author: *Jesús Abraham, Castorena-Peña* / **ORC ID:** 0000-0002-8833-1159, **CVU CONACYT ID:** 411532

ID 1st Coauthor: *Alicia Elena, Silva-Avila* / **ORC ID:** 0000-0001-7093-9898, **CVU CONACYT ID:** 260461

ID 2nd Coauthor: *Alma Jovita, Domínguez-Lugo* / **ORC ID:** 0000-0003-4988-4911, **CVU CONACYT ID:** 260410

ID 3rd Coauthor: *Jehú Efraín, Martínez-Castro* / **ORC ID:** 0000-0002-4151-947X

Abstract

In recent years, with the advent of the internet of things (IOT), computers and information technology (IT), small and medium-sized enterprises (SMEs) have been immersed in a cluster of information difficult to analyze in a conventional way which provides value that can ensure its permanence and survival in the market. Therefore, it is important to make use of data mining technology tools and statistical techniques that help analyze and better understand the information generated by sales processes, through a thorough analysis of the data. For this research, the IBM Cognos Analytic tool was used, which provides a predictive data visualization and analysis service that can be used to determine patterns, relationships, associations and meaning of a large set of data quickly and in a timely manner. Likewise, statistical techniques such as correlation and simple linear regression were used to establish the behavior of the company's product demand in the marketing, manufacturing and importation of the cleaning product sector.

Data mining, Cognos analytics, Sales process, PyMES

Resumen

En los últimos años con la llegada del internet de las cosas (IOT), las computadoras y las tecnologías de la información (TI), las pequeñas y medianas empresas (PyMES) se han visto inmersas en un cumulo de información que resulta difícil de analizar de manera convencional que proporcione algún tipo de valor que pueda asegurar su permanencia y supervivencia en el mercado. Por lo cual resulta importante hacer uso de herramientas tecnológicas de minería de datos y técnicas estadísticas que ayuden a analizar y a comprender mejor la información generada por sus procesos de venta, a través del análisis exhaustivo y minucioso de los datos. Para la presente investigación se utilizó la herramienta de IBM Cognos Analytic la cual proporciona un servicio de visualización y análisis predictivo de datos que puede ser utilizado para determinar patrones, relaciones, asociaciones y significado de un gran conjunto de datos de forma rápida y oportuna. Asimismo, se utilizaron técnicas estadísticas como correlación y regresión lineal simple para establecer el comportamiento de la demanda de productos de la compañía del sector de comercialización, fabricación e importación de productos de limpieza.

Mineria de datos, Cognos analytics, Proceso de ventas, PyMES

---

† Researcher contributing first author

Introduction

The new challenges presented to organizations today have encouraged companies to rethink organizational strategies in order to ensure their positioning and permanence in the markets (Estrada and Dutrénit, 2007). The emergence of new markets, competitors and technologies causes organizations in any industrial sector to be in a continuous search for prediction tools and techniques for data analysis that allow them to generate or sustain their competitive advantage over their rivals, so that they can improve their business processes, increase productivity and improve the quality of their products.

Greater competitiveness will only be achieved through the implementation of strategies and the use of data mining technologies, aimed at the analysis and exploitation of information generated by the processes and people of the organization (Marulanda, López and Mejía, 2017). Thus, information becomes the key element for the development and evolution of any company, since, through the timely and efficient treatment of information, it enables organizations to improve decision-making, as well as their processes, services or products and give a timely response to market demands and changes (Dueñas, 2009).

The use of strategies such as business intelligence has allowed companies to support decision-making quickly and accurately to improve performance and generate a competitive advantage for the organization (Pollo et al., 2010). According to Bravo et al. (2011), today's organizations require technological platforms that can handle large volumes of information, including its interpretation and automatic analysis to support decision-making and generate adequate production planning for companies to be more productive and efficient in their processes.

Therefore, managing the knowledge generated by the data of the different processes of the organizations will enable to know the company's resources and the behavior of its processes in such a way that it can achieve its objectives in an adequate, effective and opportune way.

Due to this, the importance of the use of data mining technological tools arises in small and medium companies so that they can compete with large companies when it comes to obtaining patterns, trends or rules that allow them to regulate their way of operating in the market (Marulanda, López and Mejía, 2017). Tolosa and Gracia (2007) state that resources such as information systems, data mining, commercial creativity and adaptation are scarce and low-priority elements in the management of SMEs.

The data mining technological tools represent for the companies a way to manage in a more efficient way their knowledge, as well as to extract interrelations and patterns of interest of a great set of data to address the development situation which the SMEs face in the market. However, there is a lack of research on analyzing the sales process of SMEs in the sector of marketing, manufacturing and import of cleaning products through data mining technologies. Therefore, the objective of this research was focused on implementing data mining technology tools and statistical techniques to find patterns and behavior of the sales process of a company in the sector of marketing, manufacturing and import of cleaning products, to support decision-making and improve the quality of service provided.

Problem

An SME, which is mainly dedicated to the commercialization, manufacture and import of cleaning products, does not have technological tools for the analysis of data, nor a structured and reliable data model that allows an analysis and exploitation of information, to establish patterns, associations and sales projections, which provide consistent information that supports the analysis and decision-making in a reliable and timely manner. Likewise, a detailed analysis of the sales process identified that the company has several areas of opportunity, for example: lack of integrity of the data collected, lack of statistics and indicators, and inefficient management of the data.

Technological tools for data analysis in the market

There are several commercial technological products available on the net for data analysis, among which IBM Cognos Analytics stands out, which is a tool that incorporates artificial intelligence techniques that help to establish patterns, trends and associations on the data. It also presents automation functionalities such as automatic data preparation and automatic modeling; in this aspect, Cognos Analytic automatically selects the best regression model for data between linear, logistic, multivariate, etc. Other tools for data analysis are shown below:

–   **SPSS Modeler**: Tool that offers accurate predictive models quickly and intuitively through its visual interface, as well as detection of patterns and trends in hidden data. In addition, it allows to model the results obtained and to understand the factors that influence the analyzed variables to take advantage of the commercial opportunities and to mitigate the risks (IBM,2016).
–   **SAP Predictive Analytics**: Software that automates the processes of selection of the prediction algorithm, creation of the predictive model and validation of its effectiveness. It is simple, intuitive and quick to execute for the user who manipulates it, in addition, it allows to use it as initial model of reference up to a more detailed granularity (Guidi, Miniati, Mazzola y Ladanza, 2016).
–   **Revolution Analytics:** Resource that provides software, services and support together to make the R statistical computing environment an attractive tool not only for academia, exploration and prototyping, but also for implementation within a company that needs it for data analysis (Revolution Analytics, 2014).

For the practical purposes of this research, we decided to use Cognos Analytics because of the range of functionality provided to the user. Once the technological tool was selected, we proceeded with the extraction, transformation and data loading process in order to carry out an exhaustive and detailed analysis of the company's sales process that would allow for the establishment of patterns, associations and trends in sales behaviour and the extraction of knowledge that had not been considered before.

Methodology

Data extraction, transformation and loading process (ETL)

In order to start the process of data exploitation and analysis using data mining technology tools and statistical techniques, the extraction, transformation and loading of data into the database was carried out in order to define and select the set of data and variables to be processed, in order to obtain statistics and patterns of the behavior of the sales process. The extraction consisted of compiling all the information regarding the company's sales area, which was related to customers, suppliers, orders and articles. Subsequently, the selection of data and attributes of interest that could provide valuable information to the company was implemented through a procedure stored in MySQL. Once the information was selected and processed, the data was uploaded to the database developed for analysis and exploitation using the Watson Analytics tool and statistical techniques (Figure 1).
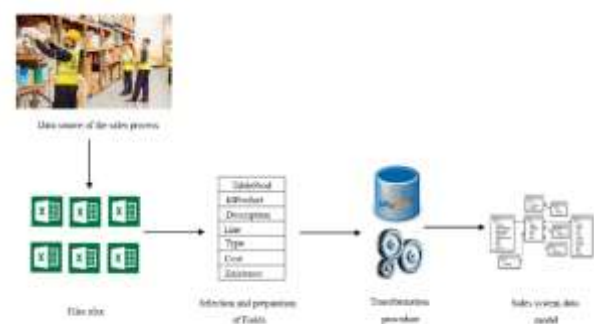

Figure 1 Data extraction, transformation and loading process (ETL)
Prepared by the authors

Implementation of Cognos Analytics in sales data analysis

Using the Cognos analytics tool, data could be analyzed and visualized quickly and clearly, discovering new patterns and trends, showing concrete results (Figure 2 and Figure 3), for example:

CASTORENA-PEÑA, Jesús Abraham, SILVA-AVILA, Alicia Elena, DOMÍNGUEZ-LUGO, Alma Jovita and MARTÍNEZ-CASTRO, Jehú Efraín. Analyzing sales to generate strategies with IBM Cognos Analytics. Journal of Business Development Strategies. 2019

– The days that report the highest sales in the company are Monday, Thursday, Tuesday and Friday with a total of 89.1%, while Sunday is the day that generates the lowest sale with respect to other days of the week. The top-selling type of product is marketing with notable displacement in Puebla followed by Veracruz, Oaxaca and Morelos. Also, the products with higher sales are ideal Venetian broom, mega economic broom, galvanized wire, parrot fan broom, Jmc clamp and fan plus broom. As for customers, it was identified that 4 customers generate the bulk of sales with 19.5% (Customer 1), 11.6% (Customer 2), 7.9% (Customer 3) and 7.2% (Customer 4).

Figure 2 Dashboard 1 Analysis of data using cognos analytics
Prepared by the authors

– It was also possible to visualize the states where there is greater presence and impact of sales, being the state of Puebla the most important with 51.6% of total sales, followed by Veracruz, Morelos and Chiapas with 13.6%, 7.2% and 6.2% respectively, as well as the analysis of the behavior of product lines in relation to the amount of sales and customer preferences. According to the data analyzed, it could be established that total sales is strongly influenced by the price of the product, city, quantity of units and the product line with a 62%, these being the variables of interest to make a purchase in the company, while the quantity of units and the price of the product as a whole have an influence of 59%.
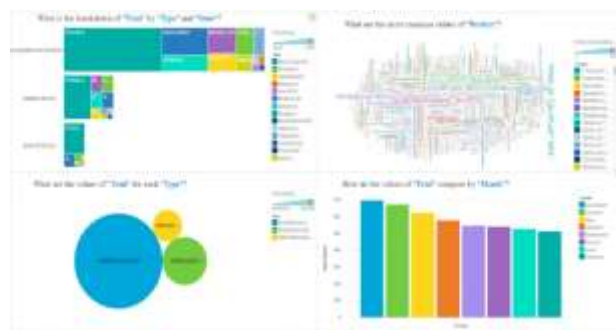

Figure 3 Dashboard 2 Analysis of data by cognos analytics
Prepared by the authors

Statistical analysis of variables of interest for sales

In order to measure the impact of the number of views the company has on monthly income, we used the correlation of variables, which showed that the monthly income from sales are not driven by the number of customers served, as can be seen in the value r=0.2131, representing almost no relationship between the variables (Table 1 and 2).

Correlation coefficient

$$r = \frac{s_{xy}}{s_x s_y} \qquad (1)$$

Nùmero de visitas (x)	Importe $ (y)
444	58036.17
320	51548.35
307	47202.16
270	30393.89
336	62552.58
303	52866.63
311	49072.34
335	54296.25
292	54894.19
246	67604.06
194	47880.83
305	69750.02

Tabla 1 Ingresos por número de visitas
Prepared by the authors

Covarianza	Desviación estándar	Coeficiente de correlación
$S_{xy} = 120752.40$	$S_x = 56.74$ $S_y = 9981.80$	$r = 0.2131$

Tabla 2 Cálculo del coeficiente de correlación
Prepared by the authors

On the other hand, it is important for the company to have a constant monitoring of the products that generate greater interest for them, which is why the linear regression technique was applied to measure and analyze the trend behavior of product IDV1, in order to plan purchases next year (2018). The construction of the prediction of units per month for the product IDV1 is presented below.

Construction of predictions

Mes	*Unidades (y)*	*t*	*ty*	*t²*
Enero	651	1	651	1
Febrero	1797	2	3594	4
Marzo	1786	3	5358	9
Abril	579	4	2316	16
Mayo	1345	5	6725	25
Junio	759	6	4554	36
Julio	1013	7	7091	49
Agosto	1044	8	8352	64
Septiembre	912	9	8208	81
Octubre	1668	10	16680	100
Noviembre	746	11	8206	121
Diciembre	1367	12	16404	144
TOTAL	13667	78	88139	650

Tabla 3 Determinación de coeficientes de unidades
Prepared by the authors

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} \qquad (2)$$

$$b = \frac{(12)(88139) - (13667)(78)}{(12)(650) - (78)^2}$$

$$b = -4.87$$

$$a = \frac{\sum y}{n} - b\frac{\sum x}{n} \qquad (3)$$

$$a = \frac{13667}{12} - (-4.87)\left(\frac{78}{12}\right)$$

$$a = 1170.58$$

$$y = a + bt \qquad (4)$$

$$y = 1170.58 - 4.87t$$

Tiempo	Año	Mes	Número de piezas
0	2017	Enero	1120
1	2017	Febrero	1123
2	2017	Marzo	1126
3	2017	Abril	1129
4	2017	Mayo	1132
5	2017	Junio	1134
6	2017	Julio	1137
7	2017	Agosto	1140
8	2017	Septiembre	1143
9	2017	Octubre	1146
10	2017	Noviembre	1149
11	2017	Diciembre	1152
12	2018	Enero	1155
13	2018	Febrero	1158
14	2018	Marzo	1161
15	2018	Abril	1164
16	2018	Mayo	1167
17	2018	Junio	1170
18	2018	Julio	1173
19	2018	Agosto	1176
20	2018	Septiembre	1179
21	2018	Octubre	1182
22	2018	Noviembre	1185
23	2018	Diciembre	1188

Tabla 4 Predicción de demanda
Prepared by the authors

Results

With the exploitation of data using Cognos Analytics data mining software, we were able to identify valuable information, since important findings unknown to the company until now appeared; some of the most relevant are:

– The type of product with the greatest impact on the company is commercialization, representing 73% of sales, followed by manufacturing representing 19% of sales and finally importing $52,394.69 representing 8% of sales.
– Four prospective customers were identified as potential customers since they bought between $8,500 and $11,500.
– 12% of revenues were from "retail" in 2,577 transactions and 88% of revenues were from "wholesale" in 1,086 transactions.
– Puebla is the state with the highest sales during fiscal year 2017 and there are seven representative states for the company, with a 94.47% of total sales.

–   There is no equilibrium in the portfolio of clients of the vendors as Employee X makes 40% of sales, Employee Y 22.5%, Employee Z 22.5% and Employee G 15%.

On the other hand, with the implementation of the statistical techniques of correlation of variables and simple linear regression, it was identified, first, that the number of views does not impact the company's income and, second, that the projection of the demand for the IDV1 product will allow the company to carry out strategies for the purchase of the product.

Recommendations

It is recommended to use this type of technological tools for the analysis of large volumes of data, due to the fact that it is an intuitive and easy tool; besides, it includes an engine for the cognitive analysis of text, where the user can ask a question in natural language to extract important data for the company; finally, it offers a series of statistical indicators on other aspects of the data analyzed.

Acknowledgments

We are grateful to the Autonomous University of Coahuila for the assistance provided for this article.

Conclusions

With the emergence of new intelligent data analysis technologies, organizations in any industry sector can reinvent themselves and rethink how they operate, in order to improve their competitiveness and productivity in the sales process. Furthermore, through the implementation of the Cognos Analytics tool, the company in question was able to determine patterns, relationships, associations and meaning of a large set of data in a fast and timely manner that would allow it to respond to the changes presented by the market and improve its decision-making in the sales process.

References

Bravo, C., Aguilar, J., Ríos, A., Aguilar, J., and Rivas, F. (2011). Arquitectura Basada en Inteligencia Artificial Distribuida para la Gerencia Integrada de Producción Industrial. *Automática e Informática Industrial, 8*, 405-417.

Dueñas, M. X. (2009). Minería de datos espaciales en búsqueda de la verdadera información. *Ingeniería y Universidad,, 13*(1), 137-156.

Estrada, S., and Dutrénit, G. (2007). Gestión de conocimiento en pymes y desempeño competitivo. *Engevista, 9*(2), 129-148.

Guidi, G., Miniati, R., Mazzola, M., & Ladanza, E. (2016). Case Study: IBM Watson Analytics Cloud Platform as Analytics-as-a-Service System for Heart Failure Early Detection. *Future Interne, 8*(3), Pp. 2-16. doi:https://doi.org/10.3390/fi8030032

IBM. (2016). *Guía de aplicaciones de IBM SPSS.* España. Retrieved in 18/11/2019, from ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/18.0/es/ModelerApplications.pdf

Marulanda , C. E., López , M., & Mejía , M. (2017). Minería de datos en gestión del conocimiento de pymes de Colombia. *Revista Virtual Universidad Católica del Norte*(5), Pp. 224-237. Retrieved from https://www.redalyc.org/pdf/1942/194250865013.pdf

Pollo, F., Amatriain, H., Rodriguez , D., Pytel, P., Ciccolella, E., Vegega, C., y otros. (2010). Ingeniería de Procesos de Explotación de Información. *Ingeniería de Software e Ingeniería del Conocimiento: Tendencias de Investigación e Innovación Tecnológica en Iberoamérica*, 52-263.

Revolution Analytics. (2014). *Revolution R Enterprise DeployR Overview Guide.* Mountain View, CA: Inc. Retrieved in 18/11/2019, from https://packages.revolutionanalytics.com/doc/7.1.0/win/DeployR_Overview_Guide.pdf

Tolosa, N & García, M. (2007). Fidelizar clientes, clave para que las pymes perduren. Universidad, Ciencia y Desarrollo. Fascículo 12: Universidad del Rosario. Retrieved from https://www.urosario.edu.co/urosario_files/4f/4f29e419-f0b0-4ee7-99bb-99a8b6f5775a.pdf