DATA AND ITS INCREASING IMPORTANCE AMONG ORGANIZATIONS AND BUSINESSES

José Miguel Orellana Parapi[1]
Edi Sunjayanto Masykuri[2]
{jose.orellana0106@gmail.com[1]}
{esunjayanto@umpwr.ac.id[2]}

Chemical faculty at University of Cuenca, Cuenca - Ecuador[1]
Automotive Technic Vocational Program Universitas Muhammadiyah Purworejo – Indonesia[2]

Abstract. Data is a driving factor in our current society, its presence is in each simple and complex system holding important insights and clues that can lead to a better performance; although its importance people usually tend to ignore its potential of discovering and organizing unusual and interesting patterns and connections within. The "Big data" era provides us with a stable and huge ground to support modern business and analytics models, which may impact directly the way how we perceive present actions and spotting possible future tendencies, improving and enhancing current decision models. This paper tries to highlight some of the various uses and strategies that several organizations are given to their databases and their data structures, and describing some of the most salient and widespread achievements made by the correct use and manipulation of a data warehouse; moreover, a revision to some syllabus of major universities and colleges were analyzed to detect tendencies on the way how professional are being formed regarding to managing data and its corresponding analysis.

Keywords: Analysis, Business, Data, Education, Models.

1. Introduction

Data is an asset that can be recollected and gathered in different ways, and from different sources [1]. Two of the most common ways of collecting data are: structured data (numerical or text inputs), and unstructured data (videos, audios, etc.).

Among the different sources that can be used to collect data are websites, forms, histories, cookies, manual inputs, gadgets, and so on [2]; moreover, many systems and algorithms have been developed for this task, such as: subscriptions, coupons, customized cards, applications, and others. In general, data can be collected from internal or external sources.

In the past, recollecting data was a difficult task, especially due to technical and physical challenges; however, recently due to new developments, and the rapid growing of the internet with a faster connection of different platforms to the cloud, this activity has become easier, partially thanks to data warehouses, which are repositories that emphasized the data collection and its storing for later use [3].

Data may come in different formats and shapes, so in order to perform the respective analyses a process of *wrangling the data* is needed [4]. Wrangling the data (or also called *data cleaning*) is a previous step of the data mining process (Figure 1) where missing or corrupted

value are removed, substituted or adapted, all of this in order to have a clean and complete database, which will be used to build the models and findings [5], the process of working with data requires to have the correct type of it due to the principle "garbage in, garbage out".

One of the most common uses that can be applied to datasets is pattern recognition, which is a technique that represent relationships among data, some of this patterns might be easily spotted by traditional methods, while other ones might be really *unpredicted discoveries* within the data [6].
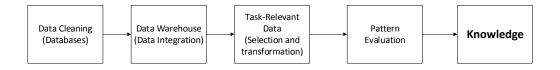

Fig. 1. Data mining process [4].

Due to the spread of new technologies, such as internet, data can be created and collected more easily; nevertheless, this increasing practice has also raised awareness about online security and privacy [7], which is especially important for companies or any other sensible organization where data is a driven factor that can be crucial for success or failure; especially, since the recent years, more and more businesses have been adapting their capabilities to a new model of business where they take decision based on models and algorithms [8].

The integration of new data processing developments and new business models has taken the current process of managing data to a new whole level called the *big data era* [9] where data can be collected, analyzed and used to construct models and systems that add value to the entire process through the creation of new insights and ideas.

2. Background and the impact of data use.

2.1. Data mining and its techniques

The impact of data in the current business model has affected the way how they create models and analyze their databases, this has enhanced some algorithms capable of spotting patterns and trends; many models have been proposed and developed throughout history, tracing back to the late 80s when the concept of data mining has been incorporated into the minds of some researches who were trying to find a good use to the data they were collecting. By the early 1980s the term data mining was born with the name of Knowledge Discovery in Data (KDD) [10], since when others contributions have been made ending up, by this time, in some widespread models for data mining activities, Table 1 shows some of the most influential techniques of data mining that are used in different sectors and fields.

Table 1. Data Mining Techniques

1	Linear regression
2	Logistic regression
3	k-nearest neighbors
4	Artificial Neural Networks

5	Deep Learning
6	Clustering
7	Support vector machine

By the use of one or more of the different machine learning tools (Table 1), many businesses have been able to realize important breakthroughs and big discoveries regarding to their business behavior than may have been ignored otherwise; these discoveries were possible thank to the capability of these models to work with big databases all at once, contrary to human intuition that has some limitations, these algorithms can use the growing potential of the computational power to run in-deep analysis in the different existing databases; besides, an integration of data from different sources or types can be made to enhance and upgrade the level of the current analysis [11], which may lead to a better understanding and comprehension of the whole picture.

Regarding to some of the important discoveries that big data has helped us to achieve we can list the followings: an improvement in bankruptcies prediction [12], prediction of customer churn in mobile services [13], enhancing the recommendation system based on past behaviors [14], market segmentations based on similar behaviors [15], and so on. All the uses of data mining have been intended to make good use of the current data to get the right information in order to take the best business decisions and actions, which allows to create a concept defined as data driven business model (DDBM) [16].

2.2. Data mining around the world

Data can be used for different purposes, such as, prediction of future trends, detection of anomalies, association of complementary behaviors and classification of similar characteristics [17], thus a specific type of knowledge has been able to be reached which is called *Business knowledge* [18], which is the state when a company gets to use their data to enhance their decisions and to support their activities through representatives models than have been created from the business itself. This practice has been increasing in its use due to an improvement in the information structure system.

Many companies are enhancing their systems with data analysis and data interpretation to offer a better user experience. Figure 2, 3 and 4 show some statistics and data taken from the webpage enlyft [19] about the global use of the oracle environment, which is one of the most common and important databases used across multiple countries and businesses [20].
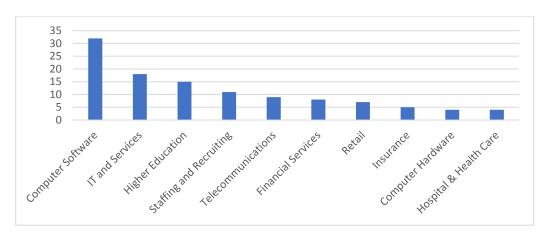
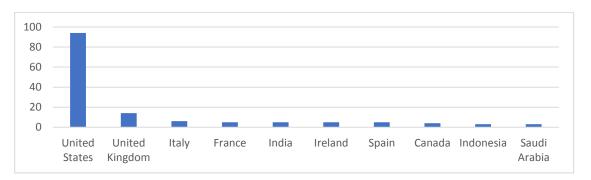**Fig. 2.** Ranking of industries that use Oracle Data Mining

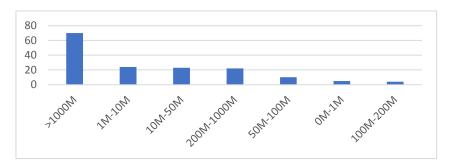**Fig. 3.** Ranking countries that use Oracle Data Mining

Fig. 4. Revenue in USD of companies that use Oracle Data Mining

Besides the widespread and intensive global use that big companies are giving to their data, there exists registers than local businesses are also using big data and machine learning concepts to boost their activities and the value they give to the marketplace; this can be made due to the versatility and the transability of some inner concepts regarding to data mining,

which has been proved to adapt to different models and conditions as long as there are enough data [21], or an adequate data pre-processing.

2.3. The incorporation of data analysis in the current education system.

Thanks to the growing awareness of new tendencies and breakthroughs many educational institutions (universities, colleges, private institutes, online platforms, etc.) have been adapting their curriculum or syllabus to capture this new field of interest. One of the most important challenges of our time is the lack of enough knowledge about the correct use and manipulation of these techniques and materials leading to a lack of professionals who know the state of the art and dominate the corresponding concepts and systems [22]. Nevertheless, recently, a growing interest in machine learning and data mining topics has been registered across different and important teaching institutions, which has been corroborated with an increasing number of enrollments in different programs related to data mining. Figure 5 shows the number of programs related to data mining that Carnegie Mellon School of Computer Science (CS - CMU) (https://www.cmu.edu) had and has to offer to its students; Figure 6 shows the number of students that have enrolled in these programs from 2010 to 2018.
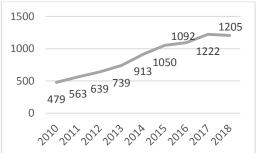

Fig. 5. Programs of CS at CMU.   **Fig. 6.** Students enrolled in SCS at CMU

Similar cases are repeated at different institutions such as: Stanford university, Georgia institute of technology, university of Washington, Colombia university and others, where an increasing rate of students pursuing data science background is demanded. Besides, the amount and extension of these programs have suffered an upgrading and have been adapted to the current times.

3. Conclusion

The increasing and spreading use and integration of new technologies and developments in the way how data is treated has allowed to realized bigger contributions and analysis to it. The organizations that use their data to run their normal actions and to create new ideas are known as data driven business models (DDBM), which has been possible thanks to a wide repertoire of machine learning tools that can be used to realize activities such as: classification, association, prediction and anomaly detection.

By using one or many of the available machine learning tools some companies or sectors have been able to create big discoveries or to create models to enhance their actions and to support their decision-making process, which allow them to keep in the market and in

business. To create important knowledge is necessary to transform correctly the available data by using the correct method, besides, it is important to deliver clearly and completely the final message and the discoveries that might have come with the analysis, this process is necessary to create *business knowledge*, which is the process of creating insights and ideas from data that can be recollected from day to day activities.

Due to the fast-changing environment that the market is facing today many companies around the world are incorporating, to their business models, new tools to keep them updated to the newest tendencies; besides, many countries have been promoting constantly these new breakthroughs allowing them to become big potencies with big economies. Also, big revenue values are related to the use of different data sources and analysis, which might be a way to ensure their market participation and income rate.

Finally, the growing demand of data users and professionals related to data mining has foster the increasing of data mining programs, especially in some prestigious institutions that currently offer more different programs with an important increasing of their enrolled students in the last years.

The way how companies manage their data is going to change the conditions of the current marketplace, which might prompt new developments and tendencies in order to get the best practices and to upgrade the existing models, which surely will represent a huge change in the near future.

References

[1]   S. D'Oca and T. Hong, "Occupancy schedules learning process through a data mining framework," *Energy Build.*, vol. 88, no. May, pp. 395–408, 2015.

[2]   P. I. Fusch and L. R. Ness, "Are we there yet? Data saturation in qualitative research," *Qual. Rep.*, vol. 20, no. 9, pp. 1408–1416, 2015.

[3]   P. Westerman, "What is Data Warehousing?," *Data Warehous.*, vol. 1, no. 4, pp. 1–30, 2012.

[4]   H. Sahu, S. Shrma, and S. Gondhalakar, "A Brief Overview on Data Mining Survey," *Ijctee*, vol. 1, no. 3, pp. 114–121, 2008.

[5]   T. Rattenbury, J. M. Hellerstein, J. Heer, S. Kandel, and C. Carreras, *Principles of Data Wrangling*. 2017.

[6]   M. Jordan, J. Kleinberg, and B. Scho, "Machine Learning & Pattern Recognition," *Zhurnal Eksp. i Teor. Fiz.*, 1937.

[7]   D. Payne, B. J. L. Landry, M. D. Dean, and D. Payne, "Data Mining and Privacy : An Initial Attempt at a Comprehensive Code of Conduct for Online Business Comprehensive Code of Conduct for Online Business," vol. 37, 2015.

[8]   P. Management, "Capturing Value from Big Data – A Taxonomy of Data-Driven Business Models Used by Start-Up Firms," 2014.

[9]   S. Kudyba, S. Kudyba, and M. Kwatinetz, "Introduction to the Big Data Era," *Big Data, Mining, Anal.*, pp. 1–16, 2014.

[10]   D. Fawzy, S. Moussa, and N. Badr, "The evolution of data mining techniques to big data analytics: An extensive study with application to renewable energy data analytics," *Asian J. Appl. Sci.*, vol. 4, no. 3, pp. 756–766, 2016.

[11]   R. P. Chakraborty, "Integration of Data Mining Systems using Sequence Process," pp. 204–211, 2015.

[12]   K. Nagaraj and A. Sridhar, "A P REDICTIVE S YSTEM F OR D ETECTION OF B ANKRUPTCY U SING M ACHINE L EARNING Kalyan Nagaraj and Amulyashree

Sridhar," vol. 5, no. 1, pp. 29–40, 2015.

[13] F. Castanedo, "<Wa_Dl (1).Pdf>," pp. 1–8.

[14] L. Kamras and W. Matslova, "An approach to a Multi-Category Recommendation System using Machine Learning Ett försök till ett Tvärkategoriskt Rekommendationssystem som använder Maskininlärning With the caveat of having limited knowledge in," 2018.

[15] R. Florez-Lopez and J. M. Ramon-Jeronimo, "Marketing segmentation through machine learning models: An approach based on customer relationship management and customer profitability accounting," *Soc. Sci. Comput. Rev.*, vol. 27, no. 1, pp. 96–117, 2009.

[16] T. Liang and Y. Liu, "Research Landscape of Business Intelligence and Big Data analytics : A bibliometrics study," *Expert Syst. Appl.*, vol. 111, no. 128, pp. 2–10, 2018.

[17] H. A. Madni, Z. Anwar, and M. A. Shah, "Data mining techniques and applications - A decade review," *ICAC 2017 - 2017 23rd IEEE Int. Conf. Autom. Comput. Addressing Glob. Challenges through Autom. Comput.*, no. September, 2017.

[18] C. Caserio and S. Trucco, "Business intelligence systems," *Contrib. to Manag. Sci.*, pp. 43–73, 2018.

[19] Enlyft, "Companies using Oracle Data Mining," 2018. [Online]. Available: https://enlyft.com/tech/products/oracle-data-mining.

[20] T. Davenport, "Oracle Data Mining 11g Release 2 Competing on In-Database Analytics," no. February, 2012.

[21] S. Y. Coleman, "Data-mining opportunities for small and medium enterprises with official statistics in the UK," *J. Off. Stat.*, vol. 32, no. 4, pp. 849–865, 2016.

[22] S. Hussain, "Survey on Current Trends and Techniques of Data Mining Research," *London J. Res. Comput. Sci. Technol.*, vol. 17, no. 1, p. 11, 2017.