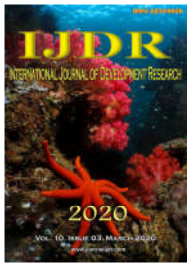
Available online at http://www.journalijdr.com

IJDR

ISSN: 2230-9926

International Journal of Development Research
Vol. 10, Issue, 03, pp. 34190-34194, March, 2020



RESEARCH ARTICLE **OPEN ACCESS**

MEASURING THE IMPACTS OF DATABASE PROCESSING UTILIZATION IN INNOVATION PROCESSES ON COMPANIES

***Vinicius Tasca Faria, Alexandre Acácio de Andrade, Jadir Perpétuo dos Santos, Júlio Francisco Blumetti Facó and Fernando Gasi**

Department of The Engineering, Modeling and Applied Social Sciences Center (CECS), Federal University of ABC, Santo André, Brazil

ARTICLE INFO

Article History:
Received 03rd December, 2019
Received in revised form
16th January, 2020
Accepted 10th February, 2020
Published online 30th March, 2020

Key Words:

Database; Innovation; Information Technology; Data Mining; Big Data.

****Corresponding author:*** **Vinicius Tasca Faria**

ABSTRACT

This work presents an empirical methodology for comprehensive measurement of innovation in companies with a database bias. The structure focuses on what is necessary to obtain a competitive advantage and the factors considered important for a company to remain active, which were found on extensive literature. This developed structure identifies, through scoring, the degree of innovation of a company, showing a correlation between data processing and the degree of innovation on the companies. To exemplify, this methodology was applied on Google, which obtained a very high score, which gives higher credibility to the methodology.

Citation: ***Vinicius Tasca Faria, Alexandre Acácio de Andrade, Jadir Perpétuo dos Santos, Júlio Francisco Blumetti Facó and Fernando Gasi*, 2020.** "Measuring the impacts of database processing utilization in innovation processes on companies", *International Journal of Development Research*, 10, (03), 34190-34194.

INTRODUCTION

The usage of information technology is increasingly present in all companies of all sizes, but the vast majority of data generated by them is overlooked (underutilized, given the potential to become information) or used for internal data generation, by means of simple statistics. By using more robust tools, it is possible to extract non-trivial information and give greater support for decision-making. Nowadays, it is very easy to generate, store, transport and transfer data. For this reason, the treatment of this data can become a great differential for any company, because the information that can be extracted has the potential to be valuable. However, without proper treatment, the information is "hidden" inside the data. They need to undergo various treatments to become significant (SATHI, 2012). Hence, how can this data be relevant to innovative processes?

Academically, the definition of innovation varies widely between authors, and can be wider, as in "making changes to something already established, introducing something new" (Oxford Dictionary, 1998, apud O'Sullivan, 2008); through definitions such as Schumpeter (1942), where innovation is a "creative destruction", that is, to generate innovation, something old must be destroyed; and by Drucker (1969), who states that "the most productive innovation is a different product or service, creating a new type of satisfaction, rather than a simple improvement"; and finally arriving at more extreme definitions, where innovation must be something completely new and not simply an "upgrade" of something already existing (Franks, 2011). Of course there are intermediate definitions, one of which is proposed by Keeley (2015), which defines some topics on what or not innovation. The first is that innovation is not equivalent to invention. The second refers to the applicability of that innovation, that is, if it will bring some return to the user (much explored by Kusiak, 2009). The third point is that little is really new, that is, all innovations are based on previously achieved knowledge (concept widely explored as "Adjacent Possible" by Johnson, 2011). The last point concerns the fact that not all innovations are product-based: some of them may be procedural innovations or the way businesses are done. It is also important to understand how the database structure works and how they are ranked. There are several ways of processing data, including statistical inference (also known as "traditional" way), data mining, big data use, machine learning, etc. (Mitchell, 1997; Witten, 2011; Franks, 2012). As shown in Figure 1, these forms of processing are interconnected (Cahoon, 2015).

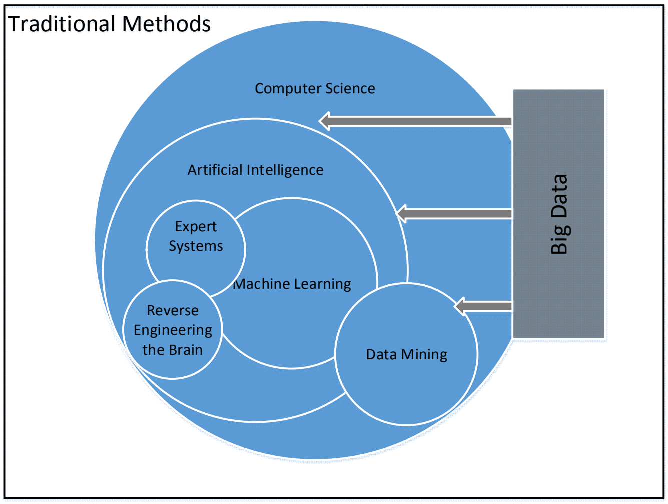


Figure 1. Relation between data processing methods. Adapted from Cahoon (2015)

METHODS

By using the data processing methods listed in Figure 1, it is possible to achieve competitive advantage (Mitchell, 1997; Witten, 2011), and this can be considered the most important factor for a company to remain active (Franks, 2012). Making only predictions about what your consumer wants (or who he is) will not bring great competitive advantage (Siegel, 2009), but this can be achieved by specific tools applied to the database mentioned above. Still according to Mitchell (1997), insight obtained by companies using any data processing (especially machine learning) is one of the interests of business and this is one of the biggest advantages of using data processing instead of classical statistics. However, it is still extremely difficult to know if the growth that the company had was due to a certain type of data processing, since there are innumerable external factors linked to the growth or decrease of a company. To try to solve this problem, a methodology is proposed that analyzes how innovative the company was in the period that it adopted some type of data processing, if it grew in that period (in relation to the value of its shares or value of gross revenue) and if it was considered innovative at the time. This methodology seeks to find a correlation between the complexity of data processing that the company uses and the degree of innovation achieved by it. To determine the level of innovation achieved, several authors were assessed, as shown on Chart 1 on the following section. As for the level of data processing complexity, a ranking score was created based on several authors and on Figure 1.It will be shown on the next section on Chart 2. As a comparison, a model on which tools can be used in each innovation situation was proposed by Kusiak (2009) so that the output of each prediction phase is the input of the next phase, but no reference was found that tried to measure directly this correlation.

RESULTS

In order to find a relationship between the degree of innovation of the studied companies and the types of data processing used in each case, a framework was developed to be applied on each of the cases. Scores were defined with different weights for each of the described criteria. All the metrics described are based on different Innovation Metrics, but the weights assigned to each of the criterion were chosen according to seminal authors in each of the areas and empirical observations of the authors of this work. This information can be found in the Table 2. The maximum theoretical score is 29 points, disregarding the possible points to be attributed by indirect techniques. Nevertheless, such a high score is highly unlikely because the company would have to have the 10 types of innovation proposed by Keeley (2015), have 100% of the departments with some focus on R&D, have a high number of patents and also be considered extremely innovative by the media and specialized rankings.

Chart 1. Score table to rank the innovation degree of the for the framework. Created by the authors

Category	Feature	Points	Main Authors
Traditional measures	Statistics of R&D	+1 to every 10% of R&D-focused departments	Patel & Pavitt (1995).
	Patents	+3 points above 60 patents, +2 between 30 and 59 and +1 with at least 10	Scherer (1965), Pavitt (1982) and Griliches et al (1991).
10 types of innovation	Profit Model	+1 if present	Keeley et al (2015).
	Network	+1 if present	
	Structure	+1 if present	
	Process	+1 if present	
	Product performance	+1 if present	
	Product system	+1 if present	
	services	+1 if present	
	Channel	+1 if present	
	Brand	+1 if present	
	Customer Involvement	+1 if present	
Classification as an innovative company	BCG Ranking	+2 in the top 10; +1 to the top 50	Tidd et al (1996).
	Indexes and local news	+2 for high impact and +1 for low and medium impact	

As for the ranking regarding the complexity of data processing on each case, a non-linear pattern was created, because it is much harder to upgrade from a simpler data-processing method to a more complex one (Mitchell, 1997). Also, since it is not easy to achieve a full usage of each data processing, a dual-type score was introduced, for situations where the company uses the technique plainly or approaches that data processing level. Chart2 shows how this ranking was arranged.

Chart 2. Score table to rank the data processing complexity of the for the framework. Created by the authors

Data processing	Approaches	Plain usage
Manual process automation	+0 points	+1 points
Big Data	+1 points	+2 points
Data Mining	+1 points	+2 points
Machine Learning	+2 points	+4 points
Artificial Intelligence	+4 points	+6 points
Expert System	+6 points	+8 points
Reverse Engineering the brain	+8 points	+10 points

Similarly to the previous ranking, the degree of complexity of data processing uses an extensive scale to minimize errors. Although it is easier to spot the use or not of a certain type of processing, it is still difficult to define whether a company uses this technique plainly or not, so it was chosen to use a small difference between those two parameters. As previously discussed, if the rankings were used with primary data, a slightly greater discrepancy could be used. For this score, the maximum theoretical score is 33 points. Like in the previous ranking, it is extremely unexpected that the maximum score is reached, even more when taking into account the last level being studied. The present work seeks to explain exactly how database processing affects the processes of innovation within certain companies to be studied in situations that have actually already happened. As the conditions that occurred in that particular situation of the company are very hard to be reproduced, the strategy of case studies seems to be the most indicated. Several cases of companies that are interesting for the purposes of this work were studied, which were: Amazon, Credit Suisse Bank, Deezer, Google, Itaú-Unibanco, Netflix Prize, Original programs Netflix, UPS and Wolfram Research. The framework was applied on all of those companies, but as Google is a globally well-know company and its results stood out, it will be more detailed below.

The Framework's Application on Google: Google is a company known by the public as very innovative. It was founded Sergey Brin and Lawrence "Larry" Page in 1998 when they were both PhD students at Stanford. Its main and best known product is its online search tool, but presently has a number of tools and products. According to the company's own website, there are more than one hundred products and Google alloys brands for users and more than twenty-five business (Google, 2017a). With so many tools, it is very difficult to analyze the data processing used by the company in a unified way, but the main focus of this work is to analyze its search engine in particular, mainly because it is the market leader, with over 67% of all searches conducted in the USA in 2009 (Nielsen, 2009). When this search engine was launched, the internet was already an established tool and there were some search engines, especially Yahoo (Wall, 2017). However, Google's search tool has become much more popular thanks to Page Rank, a technology that ranks sites according to the many terms of the search. This tool is even compared with the invention of the press regarding to the distribution of information and is nowadays is considered an indispensable tool, as compared by Vise (2007) in his book "The Google Story". As discussed in the book, the companies' tools are designed to solve problems first and then it is thought of how to monetize that particular tool. A very emblematic case concerning Google is the beginning of monetization of the search engine, which generates most of Google's revenue today (Nath, 2014). Therefore, the case of the search algorithm will be studied. Despite extensive examples of data processing usage, it was not the focus of this work to study all the various tools of the same company. Google is a service company with several different types of data processing but here primarily will be studied the PageRank algorithm.

The case studied is the product for which Google is best known. The main reason Google search engine is used instead to its competitors is its gratuity to all users, but still generating revenue for the company. Of course the speed and efficiency are also key factors on the users' choice (Vise, 2007). First, the algorithm is an approximation of an expert system as most of Google's most popular tools, An example of another Expert System approximation made by Google could be Google Maps (Pejic, 2009). Explaining how Google's algorithm is not a simple task because, although quite stable now, the process has not always been the same. A simplified and intuitive way is explained by Cutts (2013) in a didactic example, showing how Google's tool combines search terms and gives a weight to them according to the number of items found in each searched page (indexed) by the tool. This example can be seen in chart3, in which a search for the terms "Civil War" is performed. Each number on the table is a website, so it can be seen that on pages 8, 22 and 68 the two terms appear, therefore most likely they would appear first on the search engine.

Chart 3. Example of Google search ranking. Adapted from Cutts (2013)

War	3	8		22	56	68	92
Civil	2	8	15	22		68	77
Both		8		22		58	

This example illustrates how the first search filter works. However, it is obvious that the presence or absence of the search terms is not the only factor used. To rank the results, the PageRank algorithm is used, which takes into account several factors, as explained in an article published by the very founders of Google (Page, 1999). The main comparison of the search engine to the real world is related to academic publications, pointing out that on the internet (even by the time of the publication of the article) it is much easier to publish any type of unverified information, while academic articles are thoroughly sorted. The algorithm created by them takes into account various factors such as the page title, the relevance of the links (how reliable are the links within the pages), the overall theme of the page, the number of visitors that site had, etc. To have a high ranking, a website must convince an important page or multiple less important pages to have a link that redirects to that page. Figure 2(A) shows in a simplified manner how the algorithm works in a weighted manner while Figure 2(B) shows it in illustrative and more technical way.

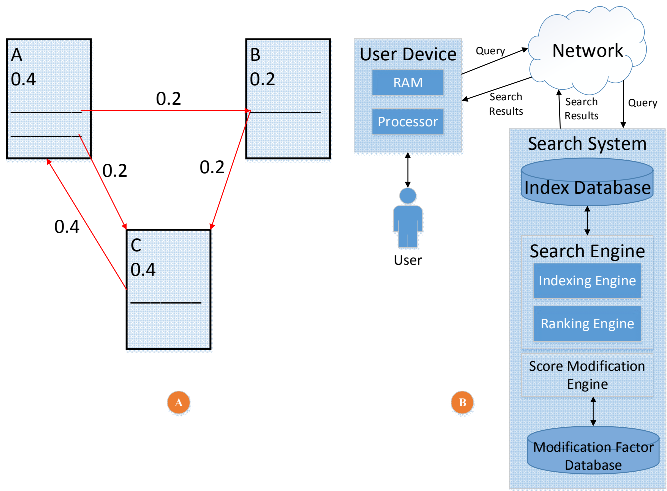


Figure 2a. Simplified representation of a weighted network. The numbers on each page are the likelihood of someone randomly go to one of them and the numbers in the rows are the weight of each vote. Adapted from Page, 1999.Figure 2(B) – How Google search works starting fromany device. Adapted from Panda & Ofitserov (2012)

Thus, it is clear how the algorithm tries to behave like an expert system by choosing pages for the user. Although there are many old details in the article of the founders of Google, more recently the algorithm also learn according to user inputs, with emphasis on the usage of mobile and tools to find low quality websites (Google, 2017b). This feature also makes the algorithm a solid machine learning tool. With the evolution of cloud processing methods, the use of Big Data is also a reality (Rijmenam, 2017; Google, 2017c).With those considerations, Chart 4 was created, reaching a score of 19 to Google on the data processing complexity criteria.

Chart 4. Scoreregarding the Data Processing complexity level for Google. Created by the authors

Google	Manual process automation	Big Data	Data Mining	Machine Learning	Artificial Intelligence	Expert System	Reverse eng the brain	Total
	Plain	Plain	Plain	Plain	Approaches	Approaches	None	
	1	2	2	4	4	6	0	19

Chart 5. Score regarding the Innovation level for Google. Created by the authors

Google	Traditional Measurements		Classification as innovative		10 types of innovation	Total
	Statistics of R&D	Patents	BCG Ranking	Indexes and local news		
	3	3	2		8	16

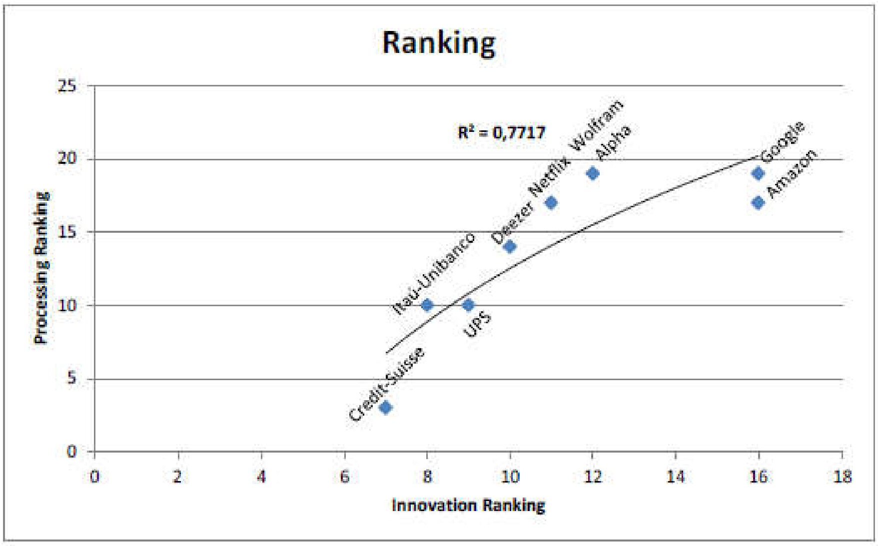


Figure 3. Correlation between of innovation level and data processing complexity degree. Created by the authors

Initially, Google was studied according to its data processing methods. Out of all the presented cases, on Google there is a strong tendency to process data with consumer insight to try to gain competitive advantage. Using traditional business growth tools and parameters, it can be seen how their growth occurred during the study period of each case Google had 23.2% average annual growth in the studied period in which the innovations took place (2005-2017). To demonstrate this amazing growth, in 2005 its revenue was $6,139 and in 2019 it was $161.857 billion (https://www.macrotrends.net/stocks/ charts/GOOG/alphabet/revenue). Finally, the company was evaluated in terms of innovation, which is shown on Chart 5. This table presents the degree of achieved innovation for the company by applying the Framework. Google has reached a total score in this criteria of 16. With Google's and other companies data, a correlation chart was created to evaluate how well a company has performed in terms of innovation and innovation level and data processing complexity degree, as can see below on Figure 3.

DISCUSSION

This work aimed to find a relationship between the degree of innovation achieved by a company and the level of data processing it has presented during a set time period.

As it was observed empirically during the course of this work, it was not simple to evidence this relationship. The degree of innovation is not easy to measure and the degree of data processing is almost always part of the main strategies of companies; therefore, it is difficult to find information on how the data is processed by researchers outside their walls.

However, when creating a methodology that intends to bypass those problems, some indications were found that this correlation actually exists. This influence and correlation between the two rankings could be better illustrated, as more case studies would be incorporated into the work. Advanced studies can benefit from the model created here and apply it to several other case studies, whether in the service sector or adapting it to different types of companies. It is also recommended that the internal staff of the companies present in the case studies check the information presented, but this can lead to a non-neutral result. This Framework could make it easier for companies to assess their level of innovation by investigating their data processing, creating a function where the only variable would be the types of data processing that each company uses. This would certainly require several tests and many more companies of each type to validate the model and be widely used. Google has shown very high indicators, with an innovation level of 16 and data processing complexity degree of 19, with the application of the developed Framework, showing that companies known to be both innovative and that uses complex data processing have high scores.

Acknowledgement: The authors are grateful to the UFABC Innovation and Sustainable Competitiveness Forum (FICS / UFABC) for intellectual support.

REFERENCES

Cahoon, J. 2015. Best of the Blogs - AI, machine learning, data mining, and big data, Automated Trader Magazine, Issue 36, Spring 2015. Available online at http://www.automatedtrader.net/articles/artificial-intelligence/153528/best-of-the-blogs_-ai-- machine-learning--data-mining--and-big--data

Cutts, M. 2013. How does Google collect and rank results?, SCIS Education Services Australia. Available online http://www2.curriculum.edu.au/scis/connections/issue_58/how_does_google_collect_and_ran k_results.html

Franks, B. 2012. Taming The Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics, ISBN: 978-1-118-20878-6, Wiley Publishing.

Giarratano, J., Riley, G. 2004. Expert Systems: Principles and Programming, ISBN-10: 0534384471, Course Technology, Fourth Edition.

Google 2017a. Our Products, Available online athttps://www.google.com/intl/en/about/products/

Google 2017b. General Guidelines Overview, Google's search quality raters guidelines. The Google Company.

Google 2017c. Soluções Para Big Data: Domine grandes volumes de dados com a nossa plataforma integrada de Big Data, Google Cloud Platform, 2017c. Available online athttps://cloud.google.com/solutions/big-data/

Google Patents. 2017 Google Patents, Engine for searching patents. Available online athttps://patents.google.com

Griliches, Z. Hall, B.H., Pakes, A. 1991. R&D, patents and market value revisited. *Economics of Innovation & New Technology Journal,* v.1, n.3.

Johnson, S. 2011. Where good ideas come from: The natural history of innovation. Penguin.

Keeley, L. et al. 2015. Ten Types of Innovation: The Discipline of Building Breakthroughs, ISBN-13: 978-1118504246, Wiley Publishing, 2013.

Kusiak, A. 2009. Innovation: A data-driven approach, International Journal of Production Economics, ISSN: 0925-5273, Vol. 122, No. 1, pp. 440–448, 2009.

Mitchell, T. M. 1997. Machine Learning, ISBN: 0070428077, McGraw-Hill Science/ Engineering/Math.

Nath, T. 2017. How The Internet Web Ad Industry Works, Investopedia.com, 2014. Available online athttps://www.investopedia.com/articles/personal-finance/121714/how-internet- web-ad-industry-works.asp

Nielsen 2009. Nielsen Online Announces March U.S. Search Share Rankings, NEW YORK– April 10. Available online athttp://www.nielsen-online.com/pr/pr_090410.pdf

Page, L., Brin, S., Motwani, R., & Winograd, T. 1999. The pagerank citation ranking: Bringing order to the web. Stanford InfoLab.

Panda, N. & Ofitserov, V. 2014. Ranking search results, US8682892 B1, Sep, 28 2012; Mar, 25.

Patel, P., Pavitt, K. 1995. Patterns of technological activity: their measurement and interpretation. In: Stoneman, Paul. Handbook of economics of innovation and technological change. Oxford Blackwell, 1995. P.14-51

Paucar, L. 2000. Inteligência Artificial, Aulas de Inteligência Artificial, Capítulo 4, UFMA, 2000, Available online athttp://www.dee.ufma.br/~lpaucar/teaching/ia2000-1/cap4.html

Pejic, A., Pletl, S., & Pejic, B. (2009, September). An expert system for tourists using Google Maps API. In 2009 7th International Symposium on Intelligent Systems and Informatics (pp. 317-322). IEEE.

Rijmenam, M. 2017. How Google Applies Big Data To Know You – Infographic, Datafloq.com, 2017. Available online athttps://datafloq.com/read/google-applies-big-data-infographic/385

Sathi, A. 2012. Big Data Analytics: Disruptive Technologies for Changing the Game, IBM Corporation.

Scherer, F.M. 1965. Firm size, market structure, opportunity and the output of patented inventions. American Economic Review, v. 55, n.4.

Siegel, E. 2013. Predictive analytics: The power to predict who will click, buy, lie, or die, ISBN-10: 1118356853, John Wiley & Sons.

Tidd, Joe; Driver, Ciaran; Saunders, Peter. Linking technological, market & financial indicators of innovation. Economics of Innovation and New Technology, v. 4, n. 3, p. 155-172, 1996.

Vise, D. 2006. The Google Story-Inside the Hottest Business, Media and Technology Success of Our Times, London: Pan Macmillan.

Wall, A. 2017. History of Search Engines: From 1945 to Google Today, Search Engine History, 2017. Available online athttp://www.searchenginehistory.com/

Witten, I.H.; Frank, E. & Hall, M.A. 2011. Data Mining - Practical Machine Learning Tools and Techniques, Elsevier, Burlington, MA 01803, USA.

****