

Challenges of External Data in Business Support: A Review

Makokha Malala Victor^{#1}, Lynet Ongachi. Sibiyia^{*2}, Ondiek J. Otieno^{#3}

^{#1} Computer Science Department, Faculty of Science, Egerton University,
P.O Box 536, Egerton, 20115, Kenya

^{#2} Computer Science Department, Faculty of Science, Egerton University,
P.O Box 536, Egerton, 20115, Kenya.

^{#3} Jomo Kenyatta University of Agriculture and Technology, Department of Computing and Information Technology, P.O Box 6200, Nairobi, 00200, Kenya

Abstract

There was and have been detectable cases far and wide where critical business choices were made dependent on investigating the accessible information. Truth be told, that immensely a larger amount of data of numerous sorts were gathered than were utilized in business knowledge, and over various years, firms have been scrambling to utilize "diagnostic instruments" so as to uncover connections and to expand firm execution i.e. increment deals, consumer loyalty, benefits, and so on and furthermore settle on crucial business choices on the constant handling. Then again, there has been a notable consideration towards information collected and used in business processes. To satisfy a business procedure, the different information objects are required and the aftereffect of an execution can be seen in creation or refresh of information. Subsequently, information is considered as the center of a business procedure in various structures and techniques. Along these lines, this paper takes a gander at the difficulties of outside information in help of business.

Keywords - Business Intelligence, data quality, real-time processing, business decisions, business analytics, and data objects

I. INTRODUCTION

There is increasing information technology support in the business world. One major support is the business decision and real-time processing in commercial and scientific applications. Such decisions could sometimes be based on mined data or data generated outside the business environment/institutions environment/ scientific application such as data from e-commerce sites, social media sites, whether, communication signals, flight data, terror attack data among others. In a research Business analytics by Saunders College of Business, enterprises were found not to make better use of the data supporting their business. Data that was slightly overtaken by time lead to "expensive" business decisions [1].

The supporting IT foundation should be fit for giving perfectly fine as hazard free choices dependent on such information. Lamentably, it is essentially

difficult to settle on hazard free choices. A portion of the dangers that are brought about are reasonable and are reliant on the sort, quality, and time of handling, sum and nature of the information that backings such a business choice. This paper attempts to look at the shortcoming of the kind of data used that may lead to a business/institution make a decision that could cost the enterprise a lot (sometimes even life). The paper also gives a recommendation of such shortcomings

II. LITERATURE REVIEW

Anne Bradstreet addresses the ethical foundations and standards for making valid business decisions. The paper is anchored on the fact that poor business decisions could lead to very big losses for any business enterprise [2]. Similarly, it can be shown that making business decisions based on low quality data can as well lead to "wrong" business decisions.

As in [3], the increasing power of computer technology that does not dispense with the need to extract meaningful information out of data sets of ever-growing size, and indeed typically exacerbates the complexity of this task. [3], further suggests that the business analytics strategy is much more dependent on the supporting amount and quality of data than the underlying mathematical and computational methods applied on the same data. Another argument comes in handy, given the recent growth in process event data; new business intelligence trends must adopt new BPA [4].

III. METHODOLOGY

Factors that affect quality of any data that is obtained outside the business environment that appear to be ignored in most cases are:

A. Time when the data was mined/obtained (Time-Value of Data)

This is a very critical component of data especially in areas such as real-time processing. Time of processing data is equally important as the time of obtaining the same data. Data of right quantity for the right application is more likely to reduce in its quality

with time difference between when it (data) was obtained and time when it was used. Another concept, *Time Value of Data – the notion that how much information is worth to the business drops quickly after it is created. The impact is even great when we move to real time analytics; an ever increasing trend. It is mathematically illustrated that time has an effect on the value of data.* For instance, we could define the importance of a data block X_j at time step t as follows:

$$AX_{j,t} = \sum_{i=1}^M w_i \cdot H(X_{j,t}|Y_{j,i}),$$

Where M is the size of the sample time window, $Y_{j,i}$ is the i^{th} sample data block and w_i is its corresponding weight. Two examples of time window are shown in Fig. 1. As we can see, w_i falls off as $Y_{j,i}$ moves away from $X_{j,t}$. Note that the normalized w_i is used, i.e.,

$$\sum_{i=1}^M w_i = 1. \text{ Finally, the importance of a time}$$

Step t is the summation of the importance values of all data blocks in t . Written in equation

$$A_t = \sum_{j=1}^N A_{X_j, t},$$

Where N is the number of data blocks at time step t [5]. This computational proof is a clear evidence of how time as a factor on data has that huge impact on the value of the data and the impact the same data can have when used to analyses and make business or enterprise decisions. There are also cases where the data value effectively drops to zero over a very short period. For these ‘perishable insights,’ if you do not act upon them immediately, you have lost the opportunity and the data becomes useless despite its quantity. The most dramatic examples are detecting faults in, say, power plants or airplanes before they explode or crash. However, many modern use cases such as fraud prevention, real-time offers, real-time resource allocation, geo-tracking and many others are also dependent on up-to-the-second data. The fast growing betting industry is a perfect example of how data mined or obtained need to be used almost as immediate as it is obtained lest to is rendered valueless. If a gambling industry provides a service such that a current team playing a matching has a probability of 0.78 winning the match, this value will drastically change if a score was made against that team. Consequently, the company is supposed to react almost immediately by changing bid prices for the same match.

B. Source of the data

A data on how clients from China consume certain commodities is likely to misguide and

business intelligence obtained from it for use in another place say India. Behaviour and environment of the clients from the two places may be completely different. This ideology in its entirety may seem less convincing but the value of data is greatly affected with the source of data- factors such as noise in the data are dependent on the source of the data. The ideal solution would collect data from all the sources (including the databases), move it into a Data Lake (a storage repository that holds a vast amount of raw data in its native format until it is needed) for historical analysis and modelling, and also provide the capabilities for real-time analysis of the data as it is moving. This would maximize the time-value of the data from both immediate and historical perspectives. Expensive read-only replicas can be used for this purpose, but a better solution is to use Change Data Capture (CDC). This effectively turns what is happening in a database into a stream of changes that can be fed continually into the Lake. CDC works against the transaction log of the database in a non-intrusive, low-impact manner.

C. Amount of data

The quantity of data may be less or too huge; “Yes too huge”; to support the best business decision from such voluminous data. The mathematical algorithms applied on the data usually assume an ideal volume of data. Exceptions quickly arise when the data is too little or too huge to make the truthiness of the guiding mathematical frameworks to manipulate the data to provide realistic results. Systems enabled by machine learning can provide customer service, manage logistics, analyze medical records, or even write a news story. Even though there are cases where When a HiPPO (highest paid person’s opinion) is in play, the decision is likely to make regardless of the quality of data or even considering its actual size in relation to the type of business support it is expected to have. With reference to the gambling industry, it will be useless to predict how a match would result in a hostile environment (whether, fans behaviour etc.) using statistics for the same match in a friendly environment

D. Noise in the Data

Data accuracy depends on the noise level of the data itself. Data with more noise is less going to be accurate. If there is real-time processing of data such as a pilot who would like to know whether to go higher (altitude), lower due to security issue, or whether conditions, any amount of noise gathered from the planes environment can have a devastating effect in the decision that would be made from it. This aspect of data may have more loud effects on applications such are military attach and terror prevention. Locating the exact location of a target especially in terror countering requires processing of camera and grid location data to be as pure as

possible as any kind of noise if not well handled may lead to completely undesired outcomes-like striking the wrong target.

IV. DISCUSSION OF FINDINGS

A basic delineation could open up a tremendous accomplishment into, what it truly takes for any sort of information regardless of the examination strategies connected on it, for it (the information) to be completely successful and solid in supporting business choices. Information assortment and culmination (wholeness) likewise have a major influence in this scene. To have a genuinely total perspective of your endeavor, you should have the capacity to investigate information from all sources, at various timescales, in a solitary place yet ensuring that every part of the information to be utilized is not too huge or excessively little to help your business choice. History has it that probably the deadliest plane accidents have occurred because of clamor in correspondence or information transmission or absence of appropriate flight following account components. This is about information and its quality in settling on essential choices. For example, the table 1 demonstrates the 10 most frightening plane smashed that have been a result of any of the three cases in flight history.

Table 1: Data about deadliest accidents in aviation history. The table excludes corporate jet and military transport accidents/hijackings (Source: Aviation Safety Network)

Year	Accidents	Casualties
1972	72	2373
1973	69	2028
1985	42	2010
1974	68	1994
1996	57	1844
1979	77	1776
1962	70	1683
1969	71	1676
1977	61	1652
1976	66	1627

In some cases, experimenting with information that is of low esteem may give a foundation what it truly takes to utilized obsolete information. It very well may be contended that the best way to comprehend

the truth about whether margarine is great or awful for you is to really comprehend what happens when you eat spread, not to keep on attempting to coax out progressively multifaceted factual relapses between well-being pointers and margarine utilization.

V. CONCLUSION AND RECOMMENDATION

Organizations today catch, dissect, translate, and utilize more information than any time in recent memory. Bosses are looking for new abilities in Data Mining, Big Data, Data Analytics, Data Science, and Business Intelligence/Analytics increasingly more habitually. The new quickly developing thought of continuous investigation that relies upon constant age of information are extremely vital in settling on any business choice that somewhat or completely relies upon information. So as to ensure that we limit any conceivable harm dependent on the nature of information used to drive towards a business choice, it is essential to consider factors, for example, time when information as acquired, commotion in the information, information linearity and measure of information to help a business objective. In conclusion, the method used to gather data determines how the data will be used, its wise enough to consider all aspects of data such as time, volume, continuity (data that is continuously streaming) before starting to obtain the data.

REFERENCES

[1] Saunders College of Business, "Business Analytics," MS in Business Analytics, p. 12, 2016.
 [2] A. B. Grinols, "The Efficacy of Teaching Ethics in Business," The Efficacy of Teaching Ethics in Business, December,2007.
 [3] M.Zanin, 2006.
 [4] A. V.-B. a. R. Colomo-Palacio, "Business Process Analytics Using a Big Data Approach," Business Process Analytics Using a Big Data Approach, p. 7.
 [5] C. Wang, "Importance-Driven Time-Varying Data Visualization," Importance-Driven Time-Varying Data Visualization, p. 8, October 2008.