International Journal of Knowledge, Innovation and Entrepreneurship
Volume 2 No. 2, 2014, pp. 82—90

The Concepts of Predictive Analytics

JAMES OGUNLEYE
Middlesex University, United Kingdom
Received 20 June 2014; received in revised form 1 November 2014; accepted 20 November 2014

ABSTRACT Predictive analytics has moved decision making process and strategic business decisions in particular significantly away from 'gut feelings' and intuition to fact- and evidence-based decision making process. Predictive analytics technology is enabling organisations to predict future outcomes and events in real time with a reasonable high level of certainty. However, much of what is known about predictive analytics or the knowledge that surrounds it—be it about the platform or application—are too often driven by software vendors who have an eye on the bottom line. This paper examines the concepts of predictive analytics, the critical elements in operationalising predictive analytics and the main issues for consideration in predictive analytics project. The paper concludes that once the predictive technology has been installed, a hypothetical testing is required on the particular data sets that would enable the organisation to develop and refine its model/s and apply it to the problem that needs to be solved. Such a step-by-step approach should help organisation to learn from any 'initial' mistakes while taking on board best practice in predictive analytics project.

Keywords: Concepts, Predictive analytics, Big Data, Environment, Architecture, Modelling, Data quality

Understanding predictive analytics

Predictive analytics is not new. The concept has been around for some time and has been used successfully by large companies—financial services and supermarket retailers in particular—operating in a small number of industrial sectors. However, the benefits and potential of predictive analytics have only recently been appreciated due largely to the phenomenon of the big data (see Ogunleye, 2013b). This new-found appreciation of predictive analytics is coupled with a desire by many corporate organisations not only to inform strategic business decisions with hard facts and evidence, but also to predict future outcomes or events with a high level of confidence.

So, what exactly is predictive analytics? Predictive analytics is the application of skills, expertise and software capabilities to extract, interrogate, analyse and transform data into clear, digestible form *feed-able* into organisation planning or decision making process. OPCC (2012, p.3) conceptualise predictive analytics as a 'general purpose analytical process that enables organisations to identify patterns in data that can be used to make predictions of various outcomes, not all of which have an impact on individuals.' Put another way, predictive analytics combines human skills and expertise with technology such as machine learning of patterns in current and historical data and the application of algorithms not only to identify patterns in the data but also to forecast future probabilities of the outcome of those patterns. Two things are clear from these conceptions of predictive analytics. First, people, tools and algorithms are critical in predictive analytics project and, second, predictive analytics is uniquely forward-looking, making predictions that based solely on data—as shown in the diagram below.

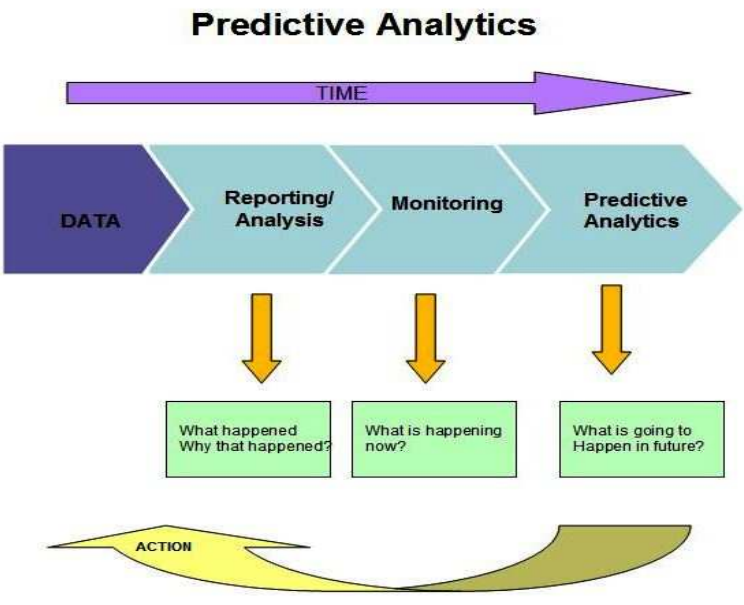

Fig 1: Predictive analytics process—source[1]

1. *www.predictiveanalyticstoday.com/what-is-predictive-analytics/*; accessed: 26.8.14

Big Data and predictive analytics

The new-found appreciation of, and interest in predictive analytics can be attributed to the phenomenon of 'Big Data', a term that describes the innovation that surrounds the possible use of digital information. The world is witnessing an explosive growth in digital and physical data and in the number of different types of datasets in the public domain and comprise of enterprise data, public data, sensor data, transactions and social media (figures 2 and 3). These datasets—85% of which are unstructured or not metric data (SAS Institute, 2012)—are huge and complex in *volume, velocity, variety, veracity and variability* they are significantly beyond the capability of standard data processing and analytic tools and, threatens traditional computing architectures. To be sure, predictive analytics is different from traditional analytics—popularly referred to as Business Intelligence. Traditional analytics is limited by the amount data and models that can be used and only supports descriptive and diagnosis analytics. Predictive analytics has the capacity to handle raw, large scale datasets and complex models (see also Ogunleye, 2013a). Predictive analytics technology is therefore critical to sense- and meaning-making of Big Data, as predictive analytics not only 'makes it possible to harness the power of big data' (Heitmueller, et al. 2014, p.1523) thereby leveraging organisation data assets, but also critical to translating Big Data into 'meaningful, useable business information' (Abbott, 2014).

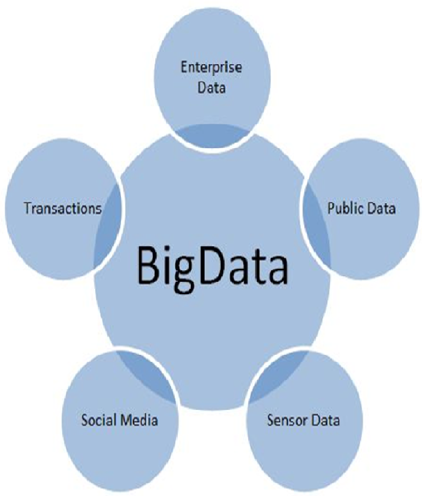

Fig. 2: Sources of Big Data—Huijbers (2012)

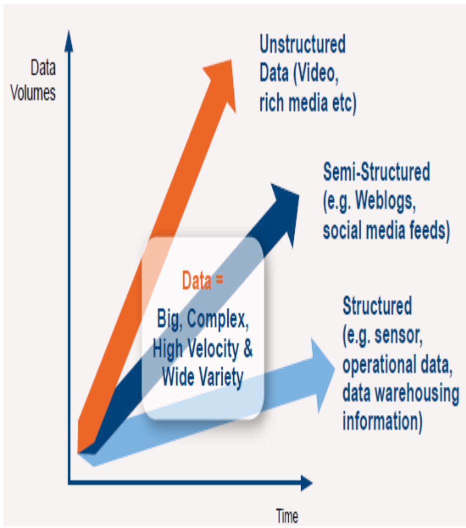

Fig 3: Types of data—IDC (2012)

As an emerging discipline and curriculum area of study, predictive analytics draws from and brings together a diverse body of knowledge such as quantitative research methods and analysis (statistics and applied maths), consumer behaviour (marketing), risk management and decision making theory (business strategy), operation research, computer science, data science, accounting and finance, etc. By drawing on such a diverse body of knowledge, predictive analytics 'solutions' can therefore be applied to and implemented in a range of sectoral contexts.

Although predictive analytics is applicable to government operations, its' use has been largely limited to the corporate sector of the economy. Take a few examples: retailers such as supermarket chains use the concept to analyse current and historical sales data, see and identify patterns in customer behaviour and use these patterns of behaviour to predict what product/s customers are most likely to buy. The use of predictive analytics can be found in commercial banks and other financial services companies, where predictive analytics is used to forecast the likelihood of a customer defaulting on a personal loan or mortgage repayments; or to determine a customer's credit card or bank overdraft limit. Heitmueller et al (2014, p.1523) reports how predictive analytics is being used to 'improve the health of patients and lower the cost of health care' in the United States. Similarly, predictive analytics technology is being used by the insurance industry in some developed economies to establish which healthcare claims are most likely to be bogus or even fraudulent. Sports might be an unusual setting for the application of predictive analytics, but the technology is being used in the United States to predict the probability that a basketball fan will renew a season ticket (see also Cao, 2012).

Operationalising predictive analytics

Predictive analytics brings together management, information technology and modelling (Miller, 2014), and three elements—environment, models and architecture—are crucial in operationalising it (Taylor, 2012). Take environment, the first element. To ensure that the right issue is addressed or the right problem is solved, there is a need for an enabling environment that encourages team work and collaboration where everyone involved subscribe to a shared goal, agree to the problem that is needed to be solved and, more importantly, take ownership of the project. This is the foundation for a cost-effective, successful application of predictive analytics in any setting.

The second element in operationalising predictive analytics is model. The word 'model' in this context refers to the relation of one set of variables to another (Miller, 2014). Modeling is at the heart of predictive analytics project and it is extremely important that an organisation has in place a process for developing predictive analytic models. According to Taylor (2012), the modelling process has to be 'repeatable, industrial-scale' to ensure effective development of 'dozens or even thousands' of required predictive analytic models.

The third element is a robust architecture for predictive analytics. The choice of architecture—be it relational or non-relational—is critical to the deployment and management of predictive analytic models in production systems. The choice of architecture will be determined by the architecture drivers and the extent to which the architecture drivers are able to handle large volume of data, process non-metric or unstructured data, reliable and tolerance of fault, provide outcome in real time, are cost–effective and quality, consistency and security-assured, etc. Nonetheless, the deployment and management of predictive analytic models should be a shared/collaborative effort that should involve a 'collaborative team of data modelers, data architects, scoring officers, and validation testers' (Chu, et al, 2007, p2).

These three elements of environment, model and architecture could be seen as both a *predictive analytics tripod* where each leg is important and a triangle where everything connects as shown in figure 3. More importantly, ensuring confidence in, and assuring the integrity of predictive analytics should be a continuous and non-stop process as shown in figure 4.

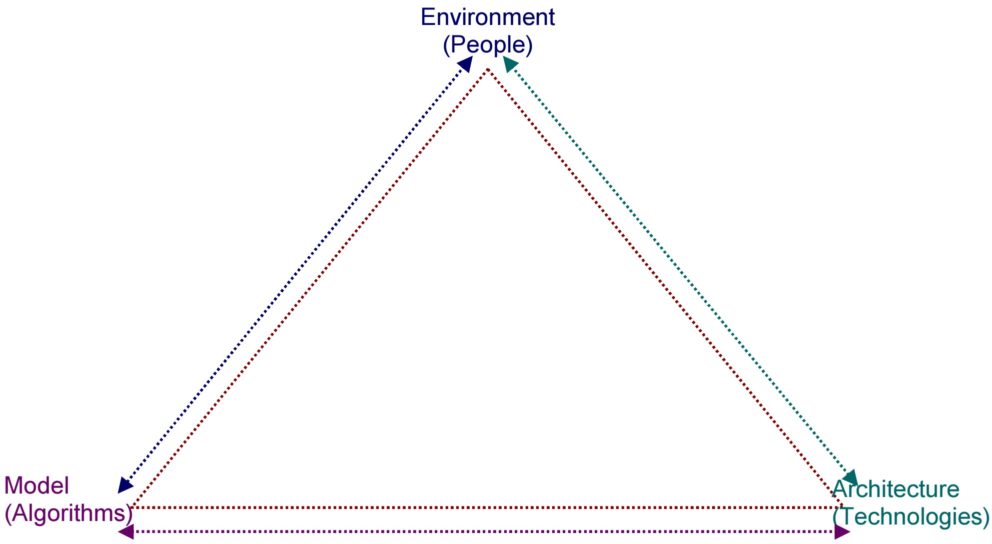

Fig. 3: Everything connects—three requisite elements in operationalising predictive analytics

Fig. 4: Assuring confidence in, and ensuring the integrity of predictive analytics is a continuous and non-stop process—Forrest Research (2013)

Issues in predictive analytics

While predictive analytics technology has made it much easier to translate big data into useful business information and revenue-generating insights, but the project might not be able to achieve desire results if it is not firmly grounded or if the project does not take into account the three elements highlighted in the preceding paragraphs**.** A central issue for consideration is the quality of data available for the predictive analytics process. Data is central to predictive analytics and it is important to have a deeper understanding of the quality of the available data before investing in predictive analytics technology (see also McCue, 2007). A recent study by Andreescu, et al. (2015, p.15) has underlined the importance of good quality of data in analytical projects. Andreescu, et al found that poor data could 'cause serious consequences for the efficiency of organizations' and that poor attention to 'quality issue could potentially lead to erroneous data mining and analysis results which in turn could lead to severe consequences, financial or otherwise.' In other words, the quality of predictive analytics 'solutions' (outcome) depends very much on the quality of the input (data). Andreescu, et al. (2015) argue the need for organisations to have a system or process in place to assure and measure the quality of their datasets. This process of assuring and measuring data quality will involve data preparation, cleaning and formatting—all of which are essential for data mining, 'the process of discovering interesting patterns and knowledge from large amounts of data' (Han, et al. 2011, p6).

Another issue for consideration in the predictive analytics project is modelling. As Taylor (2012) explains, the modelling process has to be 'repeatable, industrial-scale' to ensure effective development of 'dozens or even thousands' of required predictive analytic models—all in an attempt to search for 'meaningful relationship among models and representing those relationships in models' (Miller, 2014, p.2). So, whatever the type predictive models is deployed—be it regression or classification—an important issue for consideration is the level of user discretion is considered acceptable. User discretion, judgement and experience (or lack of it) often impact the outcome of predictive analytics. In other words, without an effective model life-cycle management build into the production system or environment, predictive analytics project might not archive a desire result (see also Chu, et el. 2007).

Legal and ethical issues are other for consideration in predictive analytics, especially where a company operates in different jurisdictions or cultures. The way information about customers are kept and mined and the 'extent to which data mining's outcomes are themselves ethical' with respect to individual customers (Johnson, 2014) should conform to the highest ethical standards. According to Schwartz (2010, p.3), it is critical that an organisation 'assess whether its decision-making with analytics reflects legal, cultural, and social norms about acceptable activities and take steps, when needed to comply with these norms.' (See also Johnson, 2013; OPCC, 2012).

Another issue for consideration is how to communicate the outcome. Data scientists and analysts might feel frustrated if they suspect a lack of seriousness on the part of business users (for example) in implementing the results of the analytics, despite the organisation's investment in the predictive analytics project (see Ogunleye, 2013b).

In other words, there might be situations where analysts feel excited about the 'insight' that their analytics have yielded, but business users, on their part, fail to convert the 'new insight' into value. It is therefore necessary that organisations have people with skills and expertise not only in analysing information but also in writing up and presenting information in clear and compelling terms. This is one way to get users to capture the benefits of predictive analytics—to get users to act on the 'new insights' yielded by the analytics 'solutions'.

Finally, an often-forgotten issue in predictive analytics is the challenge of human relations. It is important that champions of predictive analytics anticipate the human relations challenges that are often arose as a result of the introduction of predictive technology. In other words, champions of predictive analytics should be mindful of concerns by those in the organisations who might have reservations about the project. The introduction of predictive analytics technology will require attitudinal change and people in the organisation who have used to making decision based on intuition or gut feeling who consider themselves an essential part of the existing decision making process could feel that their toes are being stepped on. These individuals have an interest to protect—which is to make sure that no machine takes over their jobs! They need to be listened to, won over and assure that the predictive analytics technology is required solely as a decision-making supporting tool for the organisation.

Concluding remarks

People, tools and algorithms are at the heart of predictive analytics and key to operationalising it. As a concept, predictive analytics combines human skills and expertise with technology—machine learning of patterns in current and historical data and the application of algorithms not only to see or identify patterns in the data but also to forecast future probabilities of the outcome of those patterns. The new-found appreciation and interest in predictive analytics is attributable to the phenomenon of Big Data. As a technology predictive analytics critical to sense- and meaning-making of Big Data and makes it possible to leverage organisation data assets as well as harnessing the power of Big Data. However, an organisation needs to be clear about why it needs predictive analytics technology. Once that has been established, the next task is to state in clear and compelling terms a business proposition and the kind/s of questions the organisation seeks answers to. Besides, there are issues worth giving consideration to in order to assure the integrity of the analytics process. These issues include the quality of available data, modelling process, legal and ethical considerations, and the challenge of human relations. Once the predictive technology has been installed, it is important to do some hypothetical testing or conduct feasibility studies on the particular data sets. This is crucial in the first instance. Such testing will enable the organisation to develop and refine its model/s and apply it to the problem that needs to be solved. Such initial step-by-step approach—starting small and taking a step at a time—should help organisation to learn from any 'initial' mistakes while taking on board best practice in predictive analytics project.

Correspondence
James Ogunleye, PhD, FRSA
IWBL, Middlesex University, The Burroughs, London NW4 4BT, UK, E: J.Ogunleye@mdx.ac.uk

References

Abbott, D. (2014) Applied Predictive Analytics: Principles and Techniques for the Professional Data Analyst, New Jersey: John Wiley & Sons.

Andreescu, I. A., Anda Belciu, A., Alexandra Florea, A. and Diaconita, V. (2014) Measuring Data Quality in Analytical Projects, *Database Systems Journal*, 5 (1), pp. 15-25.

Cao, C. (2012) Sports data mining technology used in basketball outcome prediction. Masters Dissertation. Dublin Institute of Technology, Ireland. http://arrow.dit.ie/cgi/viewcontent.cgi?article=1040&context=scschcomdis; accessed: 13 April 2013

Chu, R. , Duling, D. and Thompson, W. (2007) Best Practices for Managing Predictive Models in a Production Environment, SAS Global Forum 2007, Paper 076-2007, SAS Institute Inc. Available: http://www2.sas.com/proceedings/forum2007/076-2007.pdf; accessed: 20 September 2012.

Forrest Research (2013) 'Survey: German BI market set to grow strongly'; Forrest Research Fig. August 23, 2013. http://www.automotiveit.com/survey-german-bi-market-set-to-grow-strongly/news/id-008555; accessed: 26 July 2014.

Han, J., Kamber, M., and Pei, J. (2011) *Data Mining Concepts and Techniques* (Third ed). Elsevier Inc.: p.6 and 8.

Heitmueller, A., Henderson, S., Warburton, W., Elmagarmid, A. Pentland , A. and Darzi, A. (2014) Developing Public Policy To Advance The Use Of Big Data In Health Care, *Health Aff (Millwood), 33 (9),*1523-1530.

Huijbers, C. (2012) What is Big Data? [Online] http://clinthuijbers.wordpress.com/2012/08/ 24/ what-is-big-data/; accessed: 26 September 2014.

Johnson, J. A. (2013) Ethics of Data Mining and Predictive Analytics in Higher Education, Association for Institutional Research Annual Forum, Long Beach, California, May 19-22, 2013. . Available at http://ssrn.com/abstract=2156058 or http://dx.doi.org/10.2139/ssrn.2156058; accessed: 19 June 2014.

Johnson, J. A. (2014) The Ethics of Big Data in Higher Education, International Review of Information Ethics, http://www.i-r-i-e.net/inhalt/021/IRIE-021-Johnson.pdf; accessed: 10 September 2014.

McCue, C. (2007) *Data Mining and Predictive Analysis: Intelligence Gathering and Crime Analysis*, Butterworth-Heinemann: Oxford, UK.

Miller, W. T. (2014) Modeling Techniques in Predictive Analytics: Business Problems and Solutions with R, New Jersey: Published by Pearson Education Inc.

Ogunleye, J. (2013a) Understand the Difference Business Intelligence and Business Analytics, In *IndiaBook*, available: http://www.indabook.org/d/Understanding-the-Difference-Between-Structured-and.pdf.

Ogunleye, J. (2013b) 'Before everyone goes predictive analytics ballistic', available: http://jamesogunleye.blogspot.co.uk/2013/05/before-everyone-goes-predictive.html.

OPCC (2012) The Age of Predictive Analytics: From Patterns to Predictions. Report prepared by the Research Group of the Office of the Privacy Commissioner of Canada, available: https://www.priv.gc.ca/information/research-recherche/2012/pa_201208_e.pdf; accessed: 23 July 2014

SAS Institute (2012) Big Data Meets Big Data Analytics, White Paper. SAS Institute Inc. Available: http://www.sas.com/content/dam/SAS/en_us/doc/whitepaper1/big-data-meets-big-data-analytics-105777.pdf; accessed: 2 January 2014.

Schwartz, M. P. (2010) Data Protection Law and the Ethical Use of Analytics, The Centre for Information Policy Leadership LLP [online] http://www.huntonfiles.com/files/webupload/CIPL_Ethical_Undperinnings_of_ Analytics_Paper.pdf); accessed: 26 January 2014.

Taylor, J. (2014) 'Three steps to put Predictive Analytics to Work, *Decision Management Solutions*, SaS [online] https://www.sas.com/content/dam/SAS/en_us/doc/whitepaper2/three-steps-put-predictive-analytics-work-105837.pdf; accessed: 20 August 2014.