

## **1. Contributor names and short CVs**

**Presenter:** Dr Kevin Curran, Ulster University, Northern Ireland

**Short CV:** Kevin Curran BSc (Hons), PhD, SMIEEE, FBCS CITP, SMACM, FHEA is a Reader in Computer Science at the University of Ulster and group leader for the Ambient Intelligence Research Group. His achievements include winning and managing UK & European Framework projects and Technology Transfer Schemes. Dr Curran has made significant contributions to advancing the knowledge and understanding of computer networking and systems, evidenced by over 800 published works. He is perhaps most well-known for his work on location positioning within indoor environments, pervasive computing and Internet security. His expertise has been acknowledged by invitations to present his work at international conferences, overseas universities and research laboratories. He is a regular contributor to print, online, radio & TV news on computing & security issues. He was the recipient of an Engineering and Technology Board Visiting Lectureship for Exceptional Engineers and is an IEEE Technical Expert for Internet/Security matters since 2008. He is a member of the EPSRC Peer Review College.

He is a fellow of the British Computer Society (FBCS), a senior member of the Association for Computing Machinery (SMACM), a senior member of the Institute of Electrical and Electronics Engineers (SMIEEE) and a fellow of the higher education academy (FHEA). Dr. Curran's stature and authority in the international community is demonstrated by his influence, particularly in relation to the direction of research in computer science. He has chaired sessions and participated in the organising committees for many highly-respected international conferences and workshops. He was the founding Editor in Chief of the International Journal of Ambient Computing and Intelligence and is also a member of numerous Journal Editorial boards and international conference organising committees. He has authored a number of books and is the recipient of various patents. He has served as an advisor to the British Computer Society in regard to the computer industry standards and is a member of BCS and IEEE Technology Specialist Groups and various other professional bodies.

## **2. Type of presentation proposed**

Talk

## **3. Title of the presentation**

How businesses or organizations, with the correct interpretation and visual representation can use big data to their advantage

## **4. Summary of the presentation (<100 words)**

With the recent surge in research and publication of big data over the last few years, more industries are looking to take advantage of the potential of big data. However, certain business are still struggling when it comes to exploiting this gold mine of data when it comes to gaining a competitive edge. This talk will discuss the potential difficulties of big data analysis and how correct visual representation can mean the difference between a successful big data project and an unsuccessful one and what can be done to prevent such difficulties.

## 5. Extended Abstract

*This talk outlines the potential difficulties of big data analysis and how correct visual representation can mean the difference between a successful big data project and an unsuccessful one and what can be done to prevent such difficulties. We introduce examples from the transportation industry and the data that they use in order to ensure cost efficiency and to make better future decisions for their companies. This is explained to us by how important data can be in relation to the commercial transportation services such as the trucking businesses through the use of Telematics. Telematics not only ensure driver safety through transfer of data but can actually aid efficiency and enable the company to cut down on various losses due to human error. Predictive analytics will also be explained in relation to how future 'predictions' can enable trucking companies to save the money and to make their trucks more efficient. This talk will also discuss how the Airline companies benefit from data through minimizing faults, forecasting weather in order to reduce fuel consumption etc. The various storage platforms that companies are turning to such as 'R' and Hadoop will be explained. This talk also looks at how existing businesses and organisations have used big data to make their business more efficient.*

### 5.1 Visually Representing Big Data

In terms of visually representing big data, there can be many new challenges simply because of the large volumes, different varieties and changing velocities of data that now has to be taken into account. Visualisation is critical in communicating the results and discoveries of data analytics to the decision makers. When working with big data one of the main challenges is display the results of data investigation in a way that is meaningful. This becomes more problematic when using methods such as scatterplots, which simply show the data distribution on an XY axis. If there are overlapping points on the scatterplot, which would occur with the same value appearing numerous times, none of the overlapping points are visible (SAS, 2013). This would distort the visualisation of the data distribution, yielding inaccurate results. For a small business trying to capitalise on an emerging sales opportunity, such an error could have negative outcomes and ultimately result in wasted time and effort. A possible solution to this however is data binning. This is a visualisation technique where similar data is grouped together to effectively visualise the data. With an accurate, easy to understand form of visually representing data, it can offer a shorter route to help guide decision making and become a tool to convey information critical in all data analysis. Many businesses are turning to the open source projects like "R" statistical programming language and Hadoop as to what could be a solution to their commercial needs. The link between R and Hadoop seems a reasonable one, as they are both data driven and open source. As the amount of unstructured data that is collected by organisations increases, Hadoop is growing rapidly as the main option for storing and performing operations on that data. Hadoop is a highly scalable storage platform, as it stores and distributes large data sets across a vast number of servers that operate in parallel. MapReduce refers to two separate distinct tasks that Hadoop programs perform. The first being the map job, which takes input data and processes it to produce key value pairs. The reduce job takes those key value pairs and then combines or aggregates them to produce the final results.

Hadoop can also be a saving for organisations as it is a scale out architecture that can store all of the company's data for later use at an affordable price. Which in return over a long term can save a company thousands of euros. Although it saves the company money, it is also very fast which is hugely beneficial to a company. Hadoop has an unusual storage method as it maps data wherever it is located on a cluster, with the tools for data processing are generally on the same server where the data is located, which gives us much faster data processing. The value of Hadoop in any size of business or organisation will just increase as the amount of unstructured data continues to grow (Nemschoff, 2013). R is more than a statistical package it is a programming language for doing statistics, data analysis, data mining, algorithm development, stock trading, credit risk scoring and all types of predictive analytics. It is a programming language, so you can create your own objects, functions, and packages. It has various built in and extended functions for statistical, machine learning, and visualization tasks such as, data extraction, data cleaning, and data loading as well as many others (Rickert, 2011). R is the most powerful statistical programming language, it is very flexible and it's very extensible through package mechanism.

## Conclusion

To conclude, big data is beginning to play an essential role in how businesses, large and small, approach advertising and marketing a product. Smaller businesses are finding it more difficult to adjust to big data and incorporate it into their businesses. This is somewhat because of the lack of big data solutions available for smaller companies and some smaller companies may find the task of adapting to big data to be somewhat daunting. Large online retailers are storing huge amounts of data on peoples buying habits, however finding out what to do with this data can be problematic for businesses both large and small. It does have huge potential benefits however, as data is being created at such an incredible rate companies are finding it difficult to manage and interpret this data. This data can be useless without the correct interpretation. If the data isn't correctly cleaned, using appropriate cleaning/scrubbing methods, you are left with inaccurate data which in turn may have a huge detrimental effect to a company. This becomes even more important when analysing health care data. When using a big data solution to help with patient diagnosis, incorrect or inaccurate data could lead to a wrong diagnosis and potentially put a patient's life at risk. When it comes to understanding this data and making decisions based on results from a data investigation, a correct visual representation of this data is pivotal. If the people who make the decisions can't take understand the data or interpret it incorrectly then wrong decisions will be made.

## References

- Nemschoff, M. (2013). *Big data: 5 major advantages of Hadoop*. ITProPortal. Available at: <http://www.itproportal.com/2013/12/20/big-data-5-major-advantages-of-hadoop/>
- Rickert, J. (2011). *Big Data Analysis with Revolution R Enterprise*. 1st ed. Revolution Analytics, p.Range of pages. <http://www.revolutionanalytics.com/whitepaper/big-data-analysis-revolution-r-enterprise>
- SAS. (2014). *Data Visualisation Techniques [White paper]*. <http://www.slideshare.net/AllAnalytics/data-visualization-techniques>.