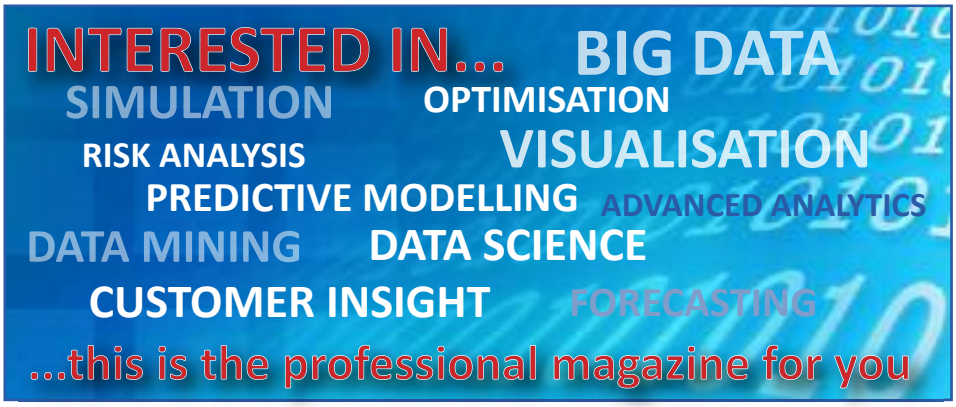


AQ
Analytics Quarterly
Spring 2013

IN THIS QUARTER'S ISSUE: Forrester Research identifies analytics leaders • Driving further and faster, in Motorsport • Analytics and Hadoop virtually inseparable by 2015 • Omnichannel, the new analytics marketing buzzword • Advanced Analytics in the palm of your hand • Predictive analytics update • Managing market uncertainty with Analytics • The Future is... • Linguistic Analytics • First data farming, now data cultivation - how do they differ? • Big Data – Moving Beyond the Hype • Artificial intelligence to make commercial sense • Boom time ahead for high end analytics professionals • Advanced analytics to ease supply chain worries • Biting back at the bookies with mathematics! • How accurate can predictions be?

THE OR SOCIETY

the Analytics Network
a section of the OR Society

Forrester Research identifies analytics leaders

Nigel Cummings

According to Forrester Research, SAS, IBM and SAP are a step above the rest in the predictive analytics world.

Forrester Research also says that SAS, IBM and SAP are leaders in big data analytics and specifically predictive analytics. Forrester's evaluation of general-purpose big data predictive analytics solution providers was obtained using the 'Forrester Wave' set of algorithms which identified companies that analyse data, possess architectures that can handle big data and use tools that span the full predictive analytics life cycle.

The report says that of ten leading analytics software vendors evaluated, SAS was clearly 'an analytics powerhouse' and, along with IBM, an 'unshakable leader' in the analysis of big data predictive analytics solutions. SAS scored highest in each of the three categories: current offerings, strategy and market presence.

Jim Davis, SAS senior vice president and chief marketing officer, said in a statement, 'Since 1976, SAS has provided enterprises with the industry's most powerful analytics, which continue to support the world's largest data sets, now called big data.' SAS is well known for providing a 'High-Performance Analytics Server, SAS Visual Analytics and SAS DataFlux Event Stream Processing Engine'.

IBM while very successful in the analytics field is a

relative newcomer having built up a portfolio of analytics enabled companies over the past few years. Unsurprisingly IBM has yet to catch up with SAS's provision of solutions and track record.

SAS's Enterprise Miner tool has become one of the most easy to learn analysis tools available, it can operate within databases or on distributed clusters to handle big data. Though according to Forrester, IBM's Smarter Planet campaign and acquisitions of SPSS, Netezza and Vivisimo represent a continued commitment to become leaders in big data predictive analytics. Forrester also says that IBM's complementary solutions, such as InfoSphere Streams and Decision Management, strengthen the appeal of their servicers for those firms that wish to integrate predictive analytics throughout their organisation.

SAP differentiates from other solution providers, like SAS and IBM, by putting its SAP HANA in-memory appliance at the centre of its consumer offering – this in-memory appliance provides users with a library of in-database predictive analytics routines. The company also offers a modelling tool that looks a lot like SAS Enterprise Miner and IBM SPSS Modeller.

A tier below the top three, Forrester listed Tibco, Oracle, StatSoft and KXEN as strong performers with unique approaches. In general, the strong performers had lower architecture scores than the leaders. Tibco's Spotfire advanced data visualisation tool offers core support for the S+ and R programming languages, which makes it attractive to data scientists. Forrester said that Oracle's analytics solution centred on offering in-database R programming combined with the strength of its in-database analytics technology.

StatSoft had a comprehensive number of analysis algorithms and was apparently very strong in manufacturing use cases, while KXEN had succeeded in 'collapsing' the normal predictive analytics life cycle by automating the predictive model discovery process; it also offered strong social network analysis.

Forrester said that Oracle's analytics solution centred on offering in-database R programming combined with the strength of its in-database analytics technology

In a tier below these were the 'contenders', which included Angoss, Revolution Analytics, and Salford Systems all of which it said had, 'a narrower focus'. Despite the narrower focus of these smaller vendors though, their customers apparently had good reason to consider them.

Forrester's research indicated that Angoss offered the best 'tooling' for decision trees and cloud solutions that firms can use to improve results quickly. Revolution Analytics on the other hand, appeared to focus more on being the de facto commercial provider of solutions based on the very popular open-source statistics programming language R; although other vendors in Forrester's evaluation offered or planned to offer R-based solutions.

Forrester's recent research clearly pointed out the leaders in the analytics marketplace, but it also indicated there were a large number of more specialised companies actively developing a presence in cloud and big data analytics solutions.

AQ

Driving further and faster, in Motorsport

Nigel Cummings

Analytics appears to be gathering momentum in velocity driven sports!

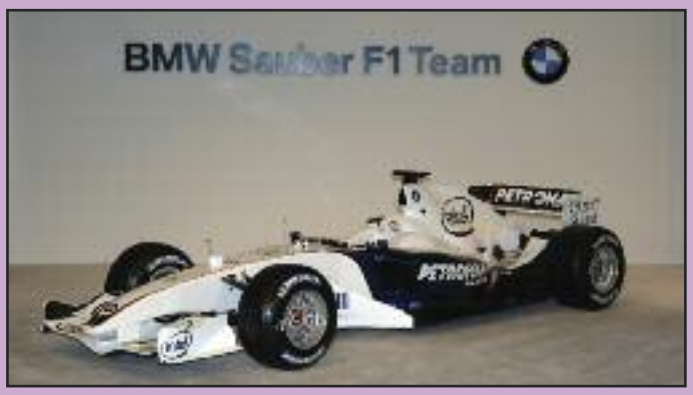
Speed combined with agility has long been a hallmark of F1 racing, so it is no surprise that advanced data analytics has been applied by some teams to exploit the new technologies available and raise performance levels.

F1 racing is a sport where one thousandth of a second can be the difference between victory and defeat. The Sauber F1 Team is amongst the first to utilise 'real-time data analysis' to help them shave those vital fractions of a second off lap times.

F1 race cars are known for their cutting edge, technology-driven designs. Drivers push cars to the limits reaching speeds over 200 mph (or 300 km/h). Those cars are also nowadays equipped with sensors, transmitters and data antennae which keeps every aspect of their performance logged. A modern F1 car is equipped with about 130 sensors, which are capable of sending enough information to fill several telephone books by the end of a two hour race. Thus it is no surprise that a resource so replete with data would eventually be subjected to analytics.

According to the Sauber F1 Team, during a typical race weekend, at least 25 GB of telemetry gets collected in total. Each of the cars running will be radioing immense streams of data back to the pits and often to the teams' headquarters. Analysis of the data—in context with the data the teams gather about their competitors—is absolutely critical to Sauber F1 team drivers getting their cars into the pole qualifying position and being first to see the chequered flag.

Every year, some 20 TB of data are collected, stored, and analysed. During the racing season which generally runs from March to November teams travel to 19 countries. Many of the teams



like Sauber F1 now rely heavily upon data analytics to constantly innovate car design and racing strategies.

Much of the data comes from wind tunnel testing and software simulations, but the critical, real-world data is derived from the weekend's practice runs, qualifying, and the race itself. Data gets used during the race, when planning for the next race, and as the team design the cars for the following season.

To give it a competitive edge, the team uses what they term 'a FlexPod ', essentially a data centre in a box. Its small form factor and low weight allows the team to take it around the world to every race; it's a reliable, mobile data centre that even operates in extreme weather conditions. The Sauber F1 Team bring their FlexPod to each race. Airfreight is a big added cost for races outside Europe, so having a small footprint, volume and weight, goes towards significantly reducing travel costs.

According to Magnus Frey, Sauber Motorsport's head of IT, the FlexPod's reliability is critical. 'Without the system at the race track running, nothing happens. We are not able to safely start the car.'

AQ

Analytics and Hadoop virtually inseparable by 2015

Nigel Cummings

Better analytics may be achieved by utilising Hadoop programming.

Government analysis departments are increasingly tapping into the power afforded by the Apache Hadoop open-source programming framework to analyse large volumes of data. This trend has not gone unnoticed by analytics software developers, who it seems, plan to incorporate purpose-built, Hadoop-based analysis functions within all future applications. Hadoop uses the MapReduce programming framework to distribute queries of large data sets across clusters of computers, which can create an efficient, cost-effective approach to analytics.

According to Gartner Research, by 2015, 65% of packaged applications with advanced analytics will come embedded with Hadoop as Hadoop-powered analytics will help reduce operational costs and IT skills requirements as well as speed up the time it takes for analysts to derive real value from data. Technology providers, in turn, will be able to deliver more task-specific analytics by utilising Hadoop directly to turn analysis into actionable insights and decisions.

Also according to Gartner, by 2015, more than 30% of all analytics projects will deliver results based on structured and unstructured data. To date, business analytics have largely been focused on tools and technologies for the analysis of structured data. This is changing as government agencies and businesses try to gain insights from new and diverse data sources such as - audio, e-mails, text, social media, video and a variety of sensors.

Then by 2016, Gartner says that 70% of leading Business Intelligence (BI) vendors will have incorporated natural-language and spoken-word capabilities. BI and analytics vendors have been slow to provide language- and voice-enabled applications to date. When they port applications to mobile and tablet devices, BI vendors tend to focus only on adapting traditional point-and-click user interfaces to touch-based interfaces.

Over the next few years though, BI vendors will enable basic voice commands as standard interfaces, followed by natural language processing of spoken or text input into SQL queries. Ultimately, personal analytic assistants will emerge that understand user context, offer two-way dialogue, and maintain conversational threads.

Where did Hadoop come from and why incorporate it with analytics?

The underlying technology was invented by Google back in their early days so they could usefully index all the rich textural and structural information they were collecting, and then present meaningful and actionable results to users. There was nothing on the market that would let them do that, so they built their own platform. Google's innovations were incorporated into Nutch, an open source project, and Hadoop was later spun-off from that.

Hadoop was designed to solve problems where there is a lot of data that does not fit nicely into tables. It is for situations where you want to run analytics that are deep and computationally extensive, like clustering and targeting.

Hadoop algorithms can be applied to a number of markets. In finance, for example, Hadoop can be used to produce an accurate portfolio evaluation and risk analysis. It can also be used to deliver better search answers to your customers so they are more likely to buy the 'things' you show them.

Hadoop was designed to run on a large number of machines that do not share any memory or disks. It can share out the work across servers automatically keeping track of the data sources enabling quick recovery if a server goes off-line. Hadoop can answer complicated computational questions dealing with large quantities of data from multiple sources because it can make use of multiple processors working in parallel.

AQ

Omnichannel, the new analytics marketing buzzword

Nigel Cummings

Google have announced a free analytics tool which will ruffle the feathers of the 'paid-for' in the web analytics industry.

Omnichannel is all about knowing your customer. It attempts to keep track of every individual purchase of a customer whether this is made in-store, through mail order or electronically including any inquiries that customer has made. The aim is identify what it is that might persuade a given individual to purchase a particular item and then target that individual in such a way as to maximise the likelihood of him or her buying the product.

Universal Analytics is what Google is using to describe the changes they have made to how their free analytics tool tracks people visiting websites. This new term relates to research done by Google which revealed that people use multiple devices in a single customer journey.

Many of us receive text message or emails from companies urging us to connect with websites and gain a bargain. We often use such information, if it appeals to us, on a tablet or desktop PC to complete the journey of discovery and ultimately make a purchase. With a QR/barcode equipped scanner app one can get product availability and compare prices electronically while in the store and hence decide whether to buy now or go elsewhere.

Sending out 'choice messages' can be very hit or miss – some people will respond positively, others negatively and many will simply ignore these messages (often diverting them straight into their 'junk mail' box). Following any response through from handset to website to purchase has often eluded these companies that utilise choice message media as a means of selling their goods. New analytics technology means that companies will no longer lose track of the people. They will be able to tell whether the person has researched the product and whether they have bought it (no matter by what means).

Whether we like it or not, we are now in the age of convergence analytics. There will still be some data challenges, but if you can encourage your visitors to sign in to your site, you are going to have access to a free tool that can track your customer's behaviour using multiple devices. This is the age of the Omnichannel and 2013 is likely to be the year of the Omnichannel marketing buzzword.

Omnichannel represents a change in measurement protocols and a philosophy that all data is good data that could provide further commercial benefit. Google is now allowing its analytics tool users to merge CRM data into Google Analytics. As a tool for analysis this represents a considerable step forward, as now extra data can be pulled into the system. Using such data in this way might allow businesses to target content for website visitors based on all purchases previously made via whatever medium.

The data-mining possibilities have just expanded exponentially. If you are a good analyst, you could have a lot of fun optimising advertising spend across all channels from brand campaigns through to retargeting. Previously this functionality was only available in Google's paid-for packages. Because this technology has been rolled out free of charge by Google, we can all build predictive models (if we want to).

This is a remarkable piece of free software and must be causing their competitors enormous headaches in trying to keep up and in trying to find ways of persuading people to spend money on their offerings – how do you compete with someone who doesn't charge!

More information can be obtained from:
www.google.co.uk/analytics/

AQ

Advanced Analytics in the palm of your hand

Nigel Cummings

Analytics from the palm of your hand is now available to IOS based tablet, android and smart phone users.

It was inevitable that analytics applications would find their way to tablet computer users, they began appearing in 2011, and the year it became possible to run some forms of Google Analytics on IOS based tablet devices. Initially quite simple to use, the saving grace of these 'early' analytics apps was that they were free.

Analytics for iPad for example was, and remains, free to download, install and use. A range of paid for analytics tools is rapidly becoming available to IOS and other operating system users. The free Google Analytics application is quite powerful insomuch as it gives users access to all Google Analytics reports, albeit with 'small advertisement' intrusions along the way. These can be removed if you pay to upgrade to the premium version. Please note that in order to harness the power of Google analytics on your iPad, iPhone or Android device you will need a Google Analytics account which you can get by registering at: http://www.google.com/analytics/

Free Google analytics for iPad for example will enable the user full access to all Google analytics reports including real-time reports and the ability to save reports for off-line access. It also provides multiple websites and network support, customisable timeframes, the ability to compare to timeframes, to each other and quick data selection: today's reports, yesterday's reports, this week's reports, this month's reports and more. In addition, there is a built-in browser to quickly check websites under scrutiny and 'Drop box' connectivity for off-line reports. The program is fully able to

communicate with the outside world due to its ability to send e-mails as Adobe Acrobat files, CSV or Microsoft Excel format files. Analytics for iPad and other tablets is optimised for daily use; it saves your last selections and presents them on app launch. It uses bookmarks to get reports; one tap of the screen is all that is needed to share reports without actually leaving the analytics app.

Chaikin Analytics LLC, a provider of institutional-grade stock analytics to asset managers and self-directed investors, have just launched Chaikin Analytics for iPad, an investment analytics workstation built specifically for the iPad and iPad Mini.

Chaikin Analytics features proprietary analytics, ratings, technical indicators and models to provide dynamically updated stock analysis, actionable trading signals, customised watch lists and flip charts with indicators for investment professionals who require a comprehensive financial analytics workstation anytime, anywhere.

The application was developed by Marc Chaikin and the team that pioneered the first real-time analytics workstation for institutional money managers and trading desks (namely, Instinet Research and Analytics, or 'R&A'), the centrepiece of the analytics is the Chaikin Power Gauge rating, a 20-factor rating that has been independently back-tested and proven successful at identifying a stock's potential 3-6 months out.

Chaikin Analytics for iPad also features buy/sell alerts and proprietary Chaikin technical indicators, including industry standard Chaikin Money Flow and Chaikin Bands. It also gives access to on-demand four-page research reports on 5,000 stocks, updated daily. In combination, these features enable the analysis needed to manage a portfolio of any size while making profitable decisions.

Chaikin Analytics for iPad is the latest addition to a suite of established stock investment tools developed by Marc Chaikin and customised for use on desktops, iOS and now Android platforms. It will cost you at least £95 per month on a rolling tariff. For more information on Chaikin Analytics go to: http://www.chaikinanalytics.com/

Another company in the premium analytics market for tablet computers is Localytics, this company's mobile app analytics gives powerful, actionable insights and deep user engagement data and is said to utilise an intuitive and powerful user interface, which provides accurate real-time user tracking, in-app funnels, and cohort analysis. Localytics support a wide range of hand held computing platforms – currently Android, iOS, HTML5, Windows Phone and BlackBerry technology platforms. It works out slightly cheaper than Chaikin, but provides more general analytics capabilities for a subscription fee of around £60 per month. A 30-day free trial of Localytics Premium for mobile devices can be accessed at: http://www.localytics.com/premium-mobile-app-analytics/

> ***Free Google analytics for iPad for example will enable the user full access to all Google analytics reports including real-time reports and the ability to save reports for off-line access.***

In conclusion, there is a choice of suppliers of complex analytics which can be performed from hand-held devices without the constraints of super computers or desktops.

AQ

Predictive analytics update

Nigel Cummings

Predictive analytics may reveal secrets you would rather not divulge

Predictive modelling could reduce unnecessary lab tests for intensive-care patients with gastrointestinal bleeding, according to a new study published in the International Journal of Medical Informatics. Researchers scrutinised a database of 746 patients and found that predictive modelling based on 11 measurements could accurately classify more than 80% of both necessary and unnecessary lab tests. They achieved an average reduction of 50% of eight common gastrointestinal lab tests. The researchers from the Massachusetts Institute of Technology said they planned to expand their work to include a number of underlying medical conditions and additional laboratory tests.

Meanwhile in the Journal of Chemical and Information Modelling, scientists have reported that using predictive modelling to project negative side effects from prescription drugs was a viable, less-costly way to test for drug side effects than current methods.

According to a paper published in the Proceedings of the National Academy of Sciences, researchers have reported success predicting when and where flu outbreaks will peak using data from Google Flu Trends and techniques for weather forecasting.

Jim Manzi, the founder and chief of Applied Predictive Technologies (APT), recently claimed that predictive analytics could help firms to get 'about three per cent better at guessing' what a consumer would like to buy. 'Realistically, what predictive modelling does not say is: 'Now I know for a fact that you want this'. What it means is I get about three per cent better at guessing - and that is worth a lot of money'.

Its software tracks customer habits in an attempt to help retailers, banks and other organisations to predict future outcomes and alter their strategies accordingly. This trend in utilising predictive analytics to work out what customers need has also been highlighted by the New York Times, which reported that US retailer 'Target' predicted a high school girl's pregnancy by analysing her shopping habits, and then with the use of predictive analytics, offered her a selection of products that are typically bought in early pregnancy. The retailer reportedly sent out coupons for other products that she might have needed for her new baby. The girl's father, meanwhile, had not been aware - up until receiving the coupons - of his daughter's situation!

Last month, the Office of Fair Trading (OFT) launched an investigation into the extent to which

Its software tracks customer habits in an attempt to help retailers, banks and other organisations to predict future outcomes and alter their strategies accordingly.

businesses are using customer data to target consumers with personalised prices. Manzi said that if society became uncomfortable with consumer-facing businesses using predictive analytics tools to tailor personalised deals then government could legislate against it. He also said, 'All predictive modelling and analytics are doing is helping companies to get a little bit more accurate at predicting who wants what.' It will be interesting to see if any more predictive analytics faux pas leak into the press in coming months.

AQ

Managing market uncertainty with Analytics

Nigel Cummings

Analytics can be applied to increase efficiencies in electrical energy supply organisations.

IBM, one of the big names in analytics service providers, recently worked with PPB (Power Procurement Business), an electricity supplier in Northern Ireland, and part of the Viridian Group of companies, to determine its electricity and fuel hedging strategies.

A range of analytical models and tools was created by IBM Global Business Services to enable PPB to meet the requirements of its new role in the single market. One of the tools developed, was an auction management tool, this enabled PPB to find new ways to optimise the management of price- and volume-risk, minimise the chance of financial loss, and reduce the time taken to manage electricity auctions.

Power NI (formerly NIE Energy) also applied the resultant analytics based tools to help it determine appropriate electricity and fuel hedging strategies. Power NI's PPB is also responsible for purchasing power under long-term contracts from independently owned generators and it sells that power directly to a spot market pool.

The creation of a Single Electricity Market (SEM) in Ireland and Northern Ireland in late 2007, created significant changes for PPB. The business was required to sell all the power it purchased from generators into a market pool, where it would be purchased by electricity supply companies. A major challenge posed by the creation of the SEM was the management of price exposure: the price of electricity in the SEM is always volatile and hard to predict and always poses considerable risks to PPB's operation.

Concerning the application of new modelling tools and analytics to the problems posed by operating in a single market, David Macartney, Commercial Manager of Power NI, said: 'We had just under two years before the single market was established to develop a completely new range of capabilities and ways of working that would enable us to comply with all the new rules and regulations. It was a very intense time, as this was uncharted territory for all of us. We were responsible for handling a turnover of roughly half a billion pounds, and we needed to rapidly develop a host of new tools that would allow us to better manage our exposure to new market risks.'

Working with IBM Global Business Services provided the solutions. Power NI became fully equipped, as a result of the alliance, to help its businesses through the challenges posed when entering into the new electricity market. In development terms the production of the tools was relatively quick for the market in which they were intended. Nevertheless a spokesman for PPB said, '…we had a team of experienced consultants working alongside us at our offices over the two years leading up to the establishment of the all-island electricity market. This was vital in helping us to find the solutions that best fit our needs.'

A large part of the project involved designing and implementing a set of analytical models and tools, which could be used to design energy trading strategies and to enable the company to auction electricity derivatives to hedge its risk in the market. One of the first applications that IBM developed was an auction management tool, this helped PPB manage the electricity auctioning process.

Contracts under auction conditions can often be valued at millions of pounds, so it is essential that well designed analytical tools are available in order for companies utilising them, to best manage those activities which involve risk. The auction management tool helped to ensure that operational risks associated with hosting auctions were mitigated and auction participants could submit bids based on the correct information. By using the application, PPB was able to reduce the time taken to determine auction results from an hour to a few minutes!

IBM also developed a strategic analysis application and this helped PPB to analyse historical costs and revenues and forecast for the future with scenario planning. Thanks to its partnership with IBM, PPB has been operating successfully since the SEM came into operation in November 2007, and the business has been able to embed new processes into daily operations to allow it to gain best advantage from analytical tools in managing the risk, pricing and auctioning of electricity derivatives.

AQ

The Future is... Nigel Cummings

Orange juice is just an orange liquid until you apply the Black Book algorithm…

Have you noticed how good orange juice is tasting straight from the super market shelves these days – no longer bitter, watery or lacking flavour, it tastes, well of oranges really, as it should do. But ensuring that the orange juice you drink tastes consistently just how the makers want it to taste is very much down to analytics.

Coca Cola for example, which makes Simply Orange and Minute Maid, two of the most popular brands of orange juices available in Europe, has an algorithm called the 'Black Book' which allows it to achieve this high level of consistency.

To bring that 'just squeezed' taste to the masses, fruit juice processing companies have to enlist technologies including maths. Physically the process entails the 'stripping' of oxygen which strips the flavour and then the addition of artificial flavour to the juice in the form of 'flavour packs' made by flavour and fragrance companies.

Apart from the physical actions required there is also a great deal of maths going on. Fruit juice manufacturers use amazingly detailed algorithms to figure everything out, the processes involved from picking to blending to packing, to ensure that every bottle or carton of juice tastes identical to every other one.

The Coca Cola Company is a good example of analytics in juice production, the company produces its successful 'big box' OJ brand utilising algorithmic controls throughout the whole manufacturing process. The architect of Coke's OJ model is consultant Bob Cross, Revenue Analytics. He says that his Black Book algorithm 'is definitely one of the most complex applications of business analytics. It requires analysing up to 1 quintillion decision variables to consistently deliver the optimal blend, despite the whims of Mother Nature.'

At the core of Coke's plan in the U.S. is 100% not-from-concentrate OJ, for which consumers are willing to pay as much as a 25% premium. Yet producing the beverage is far more complicated than bottling soft drinks. Juice production is full of variables, from weather to regional consumer preference, and Coke is trying to manage each from fruit grove to glass

Bob Cross's Black Book algorithm has a massive amount of data within it, data on over 600 of the flavours that go into an orange in fact, detailing the acidity, sweetness, etc. of each batch of juice. Once it profiles the raw juice, it issues a recipe of how to blend the juice to get it into the perfectly consistent taste.

In peak season, roughly April to June, oranges can go from grove to glass in less than 24 hours. Fibre-optic cables keep computers at Coke's juice bottling plants in constant contact so juice is piped more efficiently. Inside the bottling plant, 'blend technicians' at a traffic control centre carry out Black Book instructions prior to bottling. The weekly recipe is tweaked constantly.

Black Book incorporates external factors such as weather patterns, expected crop yields and cost pressures to help Coke plan so that supplies will be on hand as far ahead as 15 months. Using the model, if there is a hurricane or a freeze, Coke can quickly re-plan the business in 5 or 10 minutes. It even uses satellites to monitor crops to tell farmers when to pick the fruit. It is an algorithm which considers every possible angle for orange juice, but that is not all, so successful has it been, that Black Book, or versions of it, are likely to be rolled out to any company seeking to find and maintain consistency of quality, product or service.

AQ

Linguistic Analytics

Nigel Cummings

A commercial Big Brother may be waiting for you at the end of every telephone call…

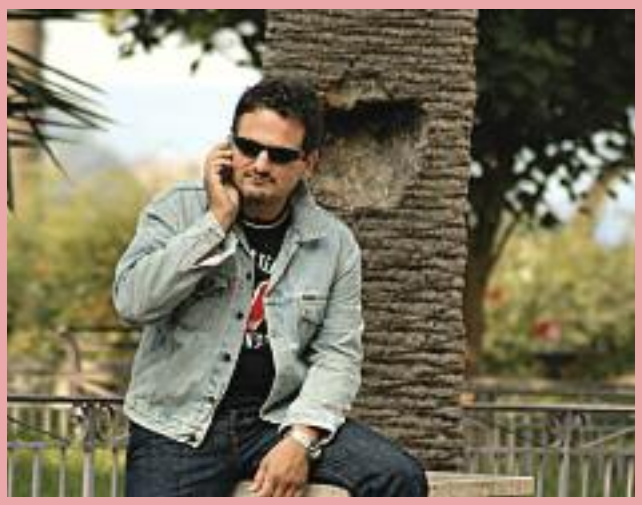
In 2009 I read a fascinating paper by Avik Sinha, a researcher at IBM, specialising in the application of natural language processing techniques for the purpose of mining information relevant to software engineering activities. His paper presented a novel linguistic engine made of configurable linguistic components for understanding natural language use case specification; and results of the first of a large scale experiment of application of linguistic techniques to industrial use cases.

Moving forward to February 2013, I now see a news report concerning Telefónica, the owner of the O2 network in the UK. Apparently Telefónica is developing technology which listens in on personal calls to draw up a psychological profile of a speaker according to their tone of voice.

A team of engineers and psychologists in Telefónica's Barcelona research facility has built a linguistic analysis engine to mine and interpret data from phone conversations. The engine can construct personality traits with 80% accuracy. The company so far has only trialled the technology using the voices of internal volunteers, though analysts briefed on the project said it could be of interest to advertisers. Telefónica insists that it has no such plans.

By tracking the location and duration of conversations, and recording spikes in activity, the 'call graph' could also alert emergency services to natural disasters or terrorist attacks. Could this be the beginnings of a telephonic 'Big Brother'?

From an analytics perspective the level of insight provided by linguistic analysis engines could be of interest to advertisers. In the UK, the company has teamed up with rival carriers Vodafone and Everything Everywhere to create an advertising sales house and virtual wallet platform. The joint venture is called 'Weve', its purpose will be to build detailed profiles of mobile phone subscribers who



opt-in to the service in exchange for promotions such as discounts from retailers, and then sell this information to advertisers looking to target specific audiences. 'Opt-in' is the operative phrase here, to utilise such information without gaining permission from communications services subscribers could be in contravention of Data Protection Act.

Apparently Telefónica's linguistic analysis engine can provide a 'deep level of granularity', for analysis. Telefónica currently says the technology will only analyse a person's tone of voice, rather than what they are saying and that it will only be used with opt-in consent from the customers. It is, however, not a giant step from this to exploitation.

A spokesman Weve has said that: 'This is a proof of concept that our innovation lab has been working on to empower people with their own data to see what it says about them. There are no plans to commercialise this and absolutely no intention of offering this information to advertisers.'

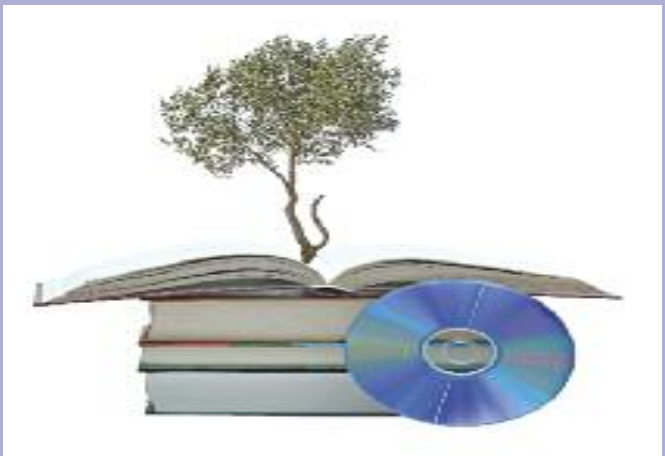
The bottom line is this technology can analyse private/confidential personal telephone calls to extract information about a caller's 'personality & preferences'. How long will it be before this information is made commercially available, despite Weve's assurances?

First data farming, now data cultivation - how do they differ?

Nigel Cummings

We recently ran a feature article on 'data farming', an emergent technology largely used by the military in the United States and highlighted by Edward Lundquist who published an article about Data Farming in Defense News.

Besides 'data farming' there is also 'data cultivation'. Data farming refers to processes and methods involved in determining the most



appropriate data collection or best fit data for data analysis. Data cultivation however, refers to a technique for capturing data about what is important in a particular event and working only with that data so that analysts' data mining models can be rerun many times to gain better results than conventional data mining would accomplish.

Data farming seems to be useful in those situations where it can enhance data that analysts already have at their disposal - the type of data that needs to be filtered in order to find the most relevant component.

Data Cultivation is already a feature of the services provided by marketing agencies. MonkeyFist Marketing, for example, provides customisation, optimisation and monetisation of integrated sales and marketing solutions for restaurants, retail stores and service-based businesses. Most marketing companies tell their clients how important data collecting is to business and how good they are at doing it. MonkeyFist however, is not like most marketing companies. They utilise a wide range of traditional and creative in-store and online data collecting programs, BUT they also utilise the relatively new technologies associated with data cultivation. They claim that data cultivation, as compared to simple data collection, is the important 1st 'Spoke' in Perpetual Cycle Results. MonkeyFist has a philosophy of data cultivation which they say has been tried tested and proven in their 'Fox In The Hen House' program. More information on this at http://monkeyfistmarketing.com/ .

Cultivation by the way is possibly going to be an important buzzword we should all be aware of. Reid Hoffman, the founder of business-oriented social networking site LinkedIn, uses it liberally with regard to the massive amounts of social data available on the Internet. He believes the availability of such data will change the business world by allowing for the development of products that yield analytics from user relationships and identities.

As long ago as June 2009, Reid Hoffman gave a video interview on IdeasProject discussing how new products will be created from social data. 'Once you have all these piles of data, people will build products out of the synthesis of this information. I think the Web 2.0 platform concerning actual identities and relationships is really just at the beginning stages of its growth- we're not close to even the midpoint of this development.' To see this interview take a look at:

http://www.websitemagazine.com/content/blogs/posts/archive/2009/06/30/online-social-data-and-relationship-cultivation.aspx

Big Data – Moving Beyond the Hype

AJ Thompson, Director of Enterprise Solutions for Northdoor

Big Data is an evolution rather than a revolution.

Hailed as the next big technology trend by vendors and analysts alike, Big Data is big news. It is also big business. With data volumes growing exponentially, statistics are continually published in the effort to convey how big is Big when it comes to Big Data. According to IDC, 2.7 zetabytes of digital data exists today and is growing at a rate of 2.5 billion gigabytes per day.

While Big Data is clearly a key business priority that is expected to add significant value to organisations, it has also become difficult to get beyond the hype.

Big Data is simply a natural progression of how organisations access, analyse and use information for the running of their businesses. Therefore we see Big Data as an evolution (rather than a revolution) that above all compels organisations to re-examine what they traditionally consider as business information.

Today businesses want relevant information at their fingertips and they want the ability to analyse this information quickly and easily. The challenge is that data has become so vast and varied that the traditional approaches to managing and analysing data can no longer meet the increasing demand. The good news is the technology is available to tackle these challenges, and Big Data tools can deliver new levels of insight fast. However knowing where to start can be overwhelming. The real key to success lies in how you go about identifying the data that will be useful and relevant to your organisation, how you examine this data, and then understanding how to store, categorise, organise and use it for competitive advantage.

There are plenty of real-world applications of Big Data today. For example in the insurance world, catastrophe and loss modelling are the two biggest data analysis challenges. Big Data is helping insurance companies better understand how events are changing and the effect this might have so that they can better manage risk. Retailers are using Big Data so that they can provide dynamic pricing and predictive analytics. They are creating up-to-the minute customer profiles which allow them to better predict buying patterns. Big brands are also using Big Data to provide better customer service. By harnessing unstructured information that sits outside the organisation, they can find out what customers are actually saying about them at any given moment and respond accordingly in real-time.

Common sense and pragmatism need to be applied when approaching a Big Data project. First clearly identify the business requirements. What are the key Big Data requirements that will provide the most value to the business? Is there a strong business case based on measurable outcomes? Is it sponsored by a business leader? Is there a pilot project that will deliver a quick win?

We recommend that those considering a Big Data initiative first focus on tangible business outcomes and then think small to think big. It might be counterintuitive but it could make all the difference in achieving successful outcomes. Here we provide practical recommendations on how to approach a Big Data project:

1. Understand that Big Data is a business-driven solution

Success will be dependent on meeting the needs of lines of business – IT is the enabler. First identify the business requirements then look to the infrastructure, data sets and analytics to support the business opportunity.

2. Establish a clear business case

For many organisations, the traditional approach to data analytics has limitations. Put a cost on it – it is the difference between having information at your fingertips in minutes as opposed to days, weeks or even months.

3. Start small, focus on quick wins

Do not try to analyse everything at once. Focus on a specific area that will deliver a quick return to demonstrate the capability of the technology.

4. Take a staged approach

Start a pilot programme by selecting a business unit or function where you think the Big Data opportunities and benefits will be. Develop proof of concepts or prototypes before you make huge technology investments. A gap analysis between your current state and desired outcome will be helpful! Where possible, benchmark yourself with industry best practice.

Artificial intelligence to make commercial sense

Nigel Cummings

Advances in AI and analytics could result in the creation of successful business products with little human intervention.

Many reports which have appeared in the popular and scientific press over the past few years regarding IBM's Watson supercomputer, a device that utilises artificial intelligence, a device seemingly imbued with the ability to beat humans at many tasks.

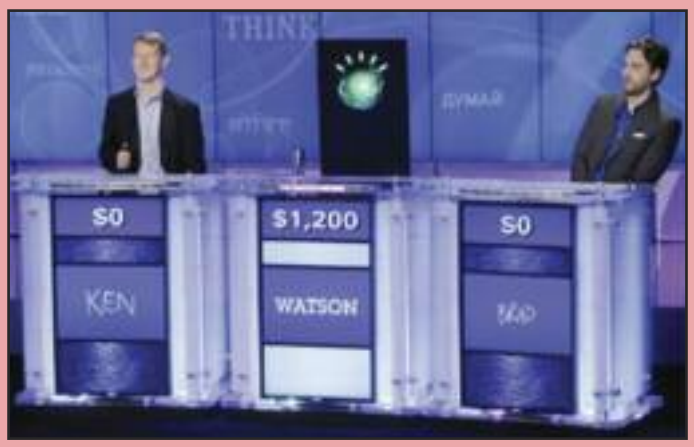
Watson hit the news originally when it was applied to game shows and in particular the popular American show 'Jeopardy'. During gameplay the supercomputer consistently outwitted human opponents both in speed and accuracy of delivery of the correct answers- its performance went a long way to proving that artificial intelligence could be useful in a real-world.

Going beyond the world of gaming though, the implications of using artificially intelligent supercomputer applications to complete tasks, which would normally require collective human intelligences is now very much a possibility.

New Watson artificial intelligence projects are in development — some on the cusp of commercialisation, others still research initiatives – all however, are at the leading edge of a much larger business for IBM and other technology companies. That market involves helping corporations, government agencies and science laboratories (for example) to find useful insights in the rising flood of data from many sources — web pages, social network messages, sensor signals, medical images, patent filings, social network activity, location data from mobile phones and others.

Advances in several computing technologies have effectively opened up this 'big data; market, and a key one of the technologies under development is the software that AI. It is no secret that IBM has been investing in AI and the big data market for years, it seems to fit in well with their analytics business which it claims has over 10,000 customers and is expected to gross $16 billion by 2015.

Watson projects are not yet big money-makers, but they probably will be soon. The projects, according to Frank Gens, Chief Analyst for IDC, infer that IBM now has the advanced technology needed and



industry expertise to do things other suppliers cannot. The new Watson offerings could affect all of us as they are services that future users will be able to tap into through a smartphone or tablet computer.

Supercomputer powered AI applications could make a significant impact in developing new business too. John Baldoni, Senior Vice-President for Technology and Science at GlaxoSmithKline, recently got in touch with IBM after seeing the terrific speed with which Watson sifted out so many wrong answers.

Sifting out wrong answers is a huge and costly challenge in drug discovery. The failure rate for new chemical compounds is high. Improving the odds of finding new drugs could increase profits and quality of life for human kind. In fact Baldoni thinks the application of AI combined with vast computing power could be key to developing safe new drugs quickly and economically.

To test this theory, Glaxo and IBM researchers recently put Watson through a test run. They fed it all the literature on malaria, known anti-malarial drugs and other chemical compounds associated with malaria treatment. Watson correctly identified known anti-malarial drugs and suggested 15 other compounds as potential drugs to combat malaria. The two companies are now discussing other projects. 'It doesn't just answer questions; it encourages you to think more widely,' said Catherine E. Peishoff, Vice-President for Computational and Structural Chemistry at Glaxo. 'It essentially says, 'Look over here, and think about this.' That's one of the exciting things about this technology.'

Watson can cook too, well it can develop new recipes for the food industry. A Watson application was recently used to develop a breakfast pastry, called a 'Spanish crescent'. The new pastry was the result of collaboration between Watson's software

Improving the odds of finding new drugs could increase profits and quality of life for human kind

developers and James Briscione, a Chef Instructor at the Institute of Culinary Education in Manhattan.

Researchers watched and talked to Briscione as he worked, selecting ingredients and building dishes, and used this data to create what can be described as an artificially intelligent virtual chef with a sound economic head for developing potentially commercially successful new food products! Watson was able to assimilate data about how the chef developed new products and to sift through combinations of around 20,000 recipes, data on the chemistry of food ingredients and ratings of flavours people like in categories like 'olfactory pleasantness'. The resultant breakfast pastry inspired by Spanish cuisine was a concoction of cocoa, saffron, black pepper, almonds and honey - but no butter!

Watson is also being used in a number of health care projects where its massively parallel processing power is of particular value.

AQ

Boom time ahead for high end analytics professionals

Nigel Cummings

According to the Harvard Business Review, the 21st century's sexiest job is the data scientist.

It suggested that such employees were 'a hybrid of data hacker, analyst, communicator and trusted adviser', but therein lies a problem; this combination of skills is extremely rare.

Many businesses are trying to recruit data analysts now that company managers recognise the benefits that big data can bring, but they are struggling to find people with the right skills.

The rush to use big data seems almost inexorable. A survey of 600 companies in the US and UK conducted in 2012 by Accenture, the management consultancy, found that two-thirds had appointed a senior figure to lead data management and analytics in the past 18 months. Even among companies that had not made such an executive appointment, 71% expected to do so in the near future.

Recruitment consultants are entering a boom time looking for data professionals because there has never been a greater demand for data analytics specialists. Cititec, which specialises in the information technology sector and has offices in London and Amsterdam, said that in the first six weeks of 2013, they received as many big data requests as they had done for the previous six months in 2012.

According to Brian McCarthy, executive director of Accenture's financial services, analytics practice in North America, 'Global demand across the industry means there is a massive shortage of data analytics skills, especially in the US and the UK'. He also says, 'Graduates with the right kinds of backgrounds for data scientist - computer science, statistics, machine learning - are coming out of the universities, but they are not coming out in sufficient numbers.'

According to Accenture many companies are turning to contractors - nearly six out of 10 companies report turning to external analysts and consultants – but they are still unable to find the people they need. The shortfall means that IT contractors who are experienced data architects or business analysts can command between £500 and £650 a day in the UK. This could be good news for O.R. and analytics graduates who are at the most highly qualified end of the profession, as PhD-level data scientists can now attract salaries of £95,000.

Accenture have published a report called 'Analytics in Action: Breakthroughs and Barriers on the Journey to ROI'. See: http://www.accenture.com/us-en/Pages/insight-analytics-action-summary.aspx for more details.

The report forecasts that in the US and UK alone, jobs demanding advanced knowledge in science, technology, engineering and mathematics (STEM) will grow five times as fast as jobs in other occupations by 2018, and four times faster than jobs in information-intensive industries such as financial services. It seems that from 2013 onwards analytics and O.R will be the careers to be in for high returns!

The report also says that emerging economies are producing STEM talent in greater numbers than developing economies, but it is not enough to meet likely demand across the globe. A survey conducted last year by Accenture's Institute for High Performance looked at demand for analytics experience in the US, China, India, UK, Japan Brazil and Singapore. It found that by 2015, with the exception of China, all of these countries would be facing a net shortage of qualified PhD graduates.

AQ

Advanced analytics to ease supply chain worries

Nigel Cummings

IBM has announced a predictive analytics system that averts supply chain disruption.

As part of their bid for worldwide domination in the field of analytics, IBM seems to be releasing new products for a new analytics markets at a fast and furious rate somewhat helped by their acquisitive policy.

The latest analytics product from their analytics arm is a predictive analytics solution designed to identify and root out problems that could lead to failure in supply chains.

Their Predictive Asset Optimisation (PAO) solution features IBM's predictive analytics software and business consulting services, it is said to be able to harness 'big data' from instrumented assets, identify irregularities in manufacturing processes, spot product irregularities and forecast a range of asset performance risks before a problem arises.

PAO utilises big data to uncover manufacturing and distribution risks and intervene before any asset disruptions might occur. The goal of PAO is to operate, maintain and manage assets throughout their lifecycle and reduce the expenses typically associated with such processes. POA is thought to be particularly useful to companies who have endured a high frequency of critical unpredicted machine failures.

Asset downtime, especially, if unplanned, is a massively expensive issue for organisations and according to IBM, the related unscheduled maintenance costs can range from three to ten times the cost of scheduled maintenance.

Decision making today will inevitably be based on the analysis of data and increasingly on the ability to apply analytics to massive data sets to extract very precise business insights. Companies realising this, and taking on board analytics at the earliest

opportunity are likely to gain considerable competitive advantage through improved business efficiency and hence create new levels of value for their customers.

IBM's PAO offering is said to be multi-adaptable across any industry in which some form of supply chain is utilised. PAO gathers facts on equipment performance, supply levels and supply volumes and anticipates potential failures in their manufacturing systems, supply chains and distribution networks before something goes wrong.

According to engineering estimates, in the U.S. alone there is a need to spend $2.2 trillion over the next five years just to bring national infrastructures up-to-date. This includes improvements to roads, bridges, water supply, sewers, electrical grids, telecommunications and more. Understanding the data about those systems, and generated by those systems, has never been more urgent.

IBM claim that their PAO system could significantly reduce these costs

AQ

Biting back at the bookies with mathematics!

Nigel Cummings

There have been many attempts to apply mathematics to horse racing – will this latest prove any more successful than its predecessors?

William Hartson, a 65 year old Cambridge University mathematician has come up with what he thinks is a sure-fire way to predict the winner of the Grand National. If successful his method could be applied to any number of races around the world – and his work could end up leaving most bookmakers with a financial headache – well maybe!

Combining sporting interests with scientific analysis to gain dividends is probably something many of wish we had the time to do. If Mr Hartson found the time, if his efforts come to fruit, we all may consider setting a little time aside for analysis of the 'Sport of Kings'. Apparently after analysing the names and ages of all of the National's previous winners this mathematician claims to have worked out a formula to pick the next champion – and this year he decided the winner would be a horse called Seabass.

By the time you read this article, the Grand National will have passed for 2013, but if his prediction was correct, and Seabass was indeed the winner, it may now be the time to consider his particular thread of sports analysis for picking out future winners.

Mr Hartson made his prediction after concluding that winners were most likely to have a name of one word consisting of eight or ten letters and beginning with the letter S, R, M or C. He also found that the winning horse was most likely to be aged nine or 10.

Seabass, an Irish 10-year-old, seemed to fit Mr Hartson's criterion well, apart from having a name falling short by one letter from Mr Hartson's ideal. Even so it scored 13 out of a possible 16 on Mr Hartson's scale, which he devised by analysing the name, first letter, number of words in the name and age of all 40 horses lining up in this year's race.

He then awarded each horse a maximum of four points for each of these criteria, depending on how similar they were to past winners from the event's 174-year history. For example, a horse whose name is eight or ten letters long – the most successful in the history of the race – would be awarded four points. But a horse with nine or twelve letters – historically less successfully – would be awarded just one point.

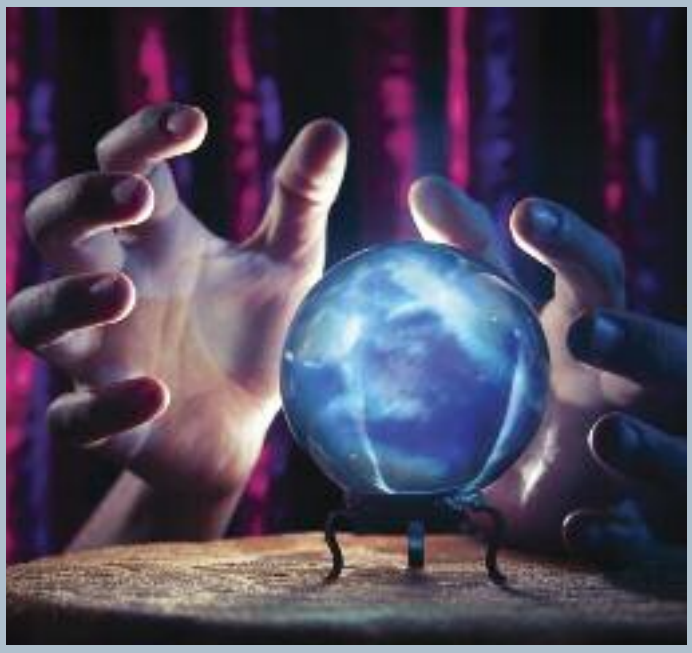
Second on his scale was Tatenen, a nine-year-old which made its debut at the National last year – it also scored 13 on the 'Hartson Index', but it was not as consistent across all four areas, while Teaforthree came out third with 12 points.

Mr Hartston, by the way is no 'flash in the pan', he is a well-known academician and author of a number of books on arithmetical coincidences and statistics. For this year's Grand National he said: 'Seabass is the only horse with consistently high scores across all four criteria. It begins with S, is a one-word name, aged ten years and has seven letters, which is only slightly short of the preferred eight. Tatenen scored an impressive 13/16 while Teaforthree scored 12/16 and shouldn't be ruled out – but their scoring pattern is less consistent.'

(Ed. With just two days to go before the race, favourite is 'On His Own', 7-1; followed by 'Seabass' at 10-1. 'Tatenen' is way down at 66-1 but 'Teaforthree' is in the top ten at 14-1. Based on name only, 'Tarquinius' might be worth a flutter at '100-1'.)

How accurate can predictions be?

Nigel Cummings and John Crocker



The answer could be very accurate indeed, so long as predictive analytics is involved…

In 1988 a reporter called Nicole Yorkin working for the Los Angeles Times, asked a team of futurologist's what life would be like in 2013. The question resulted in a number of predictions, some of which were remarkably accurate. A large number of respondents believed that in 2013 motorists would routinely use electronic navigation or mapping systems to assist them with their journeys – that prediction was quite accurate.

Many over-estimated the advances in robotics within the home, believing that by now many of us would be employing robots to carry out many of our daily needs. By contrast, none of the respondents foresaw the advances in mobile technology although many of them predicted quite accurately that systems like the Internet would be available to all and electronic mail would be in use which could allow people to send text images and even video material to each other. Another prediction which proved to be accurate was that by 2013 our children would have access to cheap but fully programmable credit card sized computers – this actually became fact with the launch of the Raspberry PI microcomputer in 2012 – a remarkable learning tool for programming and computer education costing less than £20. None predicted that data mining and analytics would be routinely used to enhance business processes. Data mining at that time was still relatively in its infancy, and the term 'analytics' was unknown!

In short, the futurologists fairly accurately predicted the development of existing technology, over-predicted our ability to develop popular science fiction gadgets and pretty much failed to think of any new ideas. To be fair, if any of them did have any new ideas, they would have been foolish to include them in their predictions unless they had already patented them.

Now, what would a similar group say if asked the same question about life in 2038? Perhaps the first major difference would be that today it would not be necessary to find individuals to ask. Today, the first thing we would do would be to use one of the many search engines to interrogate the Internet. What I really wanted to ask was what would be the future of analytics – I did not intend this query to be an ambitious one, predicting 25 years in the future could after all, be seen by many as somewhat ambitious. Instead I narrowed my search for answers to predictions about where analytics would be by the year 2020.

It did not take very long to find considerable input about the future of analytics from business

professionals making a name for themselves by publishing blogs and reports on the Internet. One report which stuck out was a piece written by Erich Siegal for the website Big Think.

By 2020 Analytics will almost certainly have eliminated car theft. The car will know enough about you to know when it is you in the driving seat. It will not let you drive the car if you are not insured for this vehicle. Your navigator will be able to perform real-time re-routing taking into account the current state of the traffic and how much charge you have left in the batteries. Although it may be capable of driving the vehicle itself, it will let you think you are in control and will only step in if you get too close to the vehicle in front or there is insufficient width for you to get through. As a price for a lower insurance premium, it will also control your speed keeping you within the limits which may be adjusted dynamically according to the time of day and the density of the traffic. Road signs may well have been eliminated from the side of the road – the ubiquitous satnav having made them redundant.

Other areas where analytics will affect our lives include menu selection. Your friendly kitchen knows what is in stock and what your preferences are so will offer you a choice of menus taking into account any special dietary needs. It will also check whether there are any items which are nearing their use by date and advise you accordingly. It may even play you a video of what to do at each stage. As you put packets and jars back, so your 'kitchen' will record what you have used and hence whether it is time to replace anything. It will draw up a list and check the prices at the supermarkets in your area.

It is hard to know what else your mobile phone will do. I would expect that it will have replaced your flexible friend and act as a debit/credit card for whatever accounts you have. It will be able to tell you exactly how much you have, or haven't, got in each account including how much is already committed via standing orders and direct debits. I suspect it will also be able to communicate with your home entertainment system whose remote control will have been replaced by voice recognition and a sensor which allows you to sit in your armchair and use your hands and fingers to do the same on the large screen as you can on the phone or iPad.

Your friendly kitchen knows what is in stock and what your preferences are so will offer you a choice of menus taking into account any special dietary needs.

There will be services similar to 'Moonpig' which will allow you to choose not just cards but gifts, have them wrapped, personalised and delivered the next day anywhere in the world all from your armchair having, of course, been reminded that it is your wife's birthday tomorrow. It will also remind you what you bought her previous years.

AI, robotics, fusion-power and endless leisure time will all still be just ten years away!

Predictive analytics it seems, will be affecting our lives substantially in our future, so much so in fact, we'll all be lost without it! Tomorrow after all, is just a day away!

AQ