school of social and behavioral sciences

DO YOU FEAR AN ANALYTICS DRIVEN COMPETITOR? YOU SHOULD!

AREA: IE SCHOOL OF SOCIAL AND BEHAVIORAL SCIENCES

PROFESSOR: **MANOEL F. A. GADI**

E-mail: mfalonso@faculty.ie.edu

Disruptive business models and disruptive technology are generating a frenzy in the race to have the best of breed technology and systems. The same is becoming reality in the field of statistical & computational models for Analytics. Companies like Capital One with its thousand simultaneous tests, Amazon with its recommendation systems and Google with its raking algorithm among others formed an army which took Analytics to a complete new level by making prescriptive modeling and testing the bedrock of its competitive advantage. In the financial services world companies such as Barclay Card, Credicard and Thinkoff CSB in the UK, Brazil and Russia had great success repeating the pioneers' strategies. This peer pressure pushed Capital One to be so knowledgeable about their customer that they achieved a point of knowing what 80% of callers wanted to discuss even before customers actually made a call to customer service. A more recent example of this greater knowledge of their users comes from Amazon. With their powerful recommendation models they are intending to ship products to potential buyers before the consumer even knows the books or goods exist and one may be surprised that Amazon is expecting very few products to be returned as unwanted. Examples in this article are mainly from for profit organizations; however, great added value of Analytics can also be seen in Wikipedia and its fundraising strategy[1], and in Bill and Melinda Gates Foundation and its Evaluation-Policy[2] which comprises of setting up competing strategies with its own goals and priorities and measuring the work results for society in a very quantitative way.

In this article we will discuss how companies blend Analytics as part of their competitive advantage to become much stronger players while including lots of interesting real world examples. We start by bringing all readers to the same page, giving some flavor of the ingredients, recipes and outputs of a model process, and showing a model in use. We then quickly expand that knowledge showing an X-RAY of a single company and how Analytics in its many forms fit every part of this company's Value Chain. We continue by showing an overview of the most preeminent models from companies recognized by their analytical skills and their success.

We conclude by introducing a new angle for looking into companies that have mastered Analytics by separating them into three groups: companies with Analytics as core competitive advantage; companies who were able to blend Analytics with their previous competitive advantage and look inside Startups that are committed to Analytics. The need for this new angle becomes evident when we look at the implementation of Analytics. First, the motivation for implementing Analytics varies largely and second, different challenges lead to specific road maps. We close this session and the article with a discussion on why, for the majority of Startups, Analytics cannot be the sole competitive advantage but the main driver for pivoting.

1 http://meta.wikimedia.org/wiki/Fundraising_2010/Banner_testing
2 http://www.gatesfoundation.org/How-We-Work/General-Information/Evaluation-Policy

Published by IE Publishing Department.
Last revised, June 2014.

1) THE MINDSET OF AN ANALYTICS DRIVEN COMPETITOR

The first problem a beginner in the field will face will be to identify the many terms referring to the qualities of an Analytics Driven Competitor. For example, Capital One [2] used the term **Information Based Strategy** (IBS) to define the involvement of the entire corporation on using verifiable facts using historical data to justify actions, test hypothesis and predict future outcomes. According to [1] Procter and Gamble used internally the term **Data-Based Decision Making** (DBDM), Thomas Davenport in the same article uses the terms **Fact Based Analysis** (FBA) and **Fact-based Decisions** (FBD) to refer to same sort actions.

Another used term more popular in the health industry is Evidence-based Decision. This term is particularly interesting and funny as it gives the opportunity to play with words, and to discuss, that Decision-based Evidence Making is not at all the same as Evidence-based Decision Making. To understand the difference here we have to center our attention on the reasons for data gathering and reporting in the first place. For instance, if data gathering and reports are being collected for an ad-hoc analysis with a much delimited remit, it is not unlikely that this data selection will have some sort of bias that is where the difference shows.

The analysis may have been requested to prove the point of a decision already taken in the head of a senior person and the data gathering and report generation is no more than "Evidence Making" to justify this decision. Evidence-based Decision Making is the opposite. Evidence triggers decisions, however, in the case a hypothesis needing testing, a very careful and fair analysis (and sometimes testing process) needs to be carried out to avoid evidence fabrication.

My particular is the term **analytics** and nothing else. If we define Analytics to be something closely related with data (with evidence), the concept would also apply to reasoning and one would agree that we do not need computers and statistical methods to make a good reasoning. We just need a good premise, a sound or grounded principle. Deductive thinking would also fall inside the Analytics remit. Therefore in this article we center Analytics to be the field of study where decisions are data driven, evidence based, where complexity of the data demands statistical or machine learning methods to arrive to conclusions or models, and where the amount of data demands computational processing power.

Although nowadays (since 2012) more people are leaving the term Analytics and starting to place everything related to information, processing, intelligence and decision making into the term **Big Data**. Using the term analytics, Thomas Davenport in [1] has an interesting 10 item list to map the DNA of a company competing on analytics. He identified that:

1. They have sophisticated information systems and rigorous analysis – Company Wide.
2. Senior executives recognize Analytics capabilities development and maintenance as a primary focus.
3. Fact-based decision is part of the culture therefore constantly emphasized and communicated by senior executives.
4. HR hires people with the very best analytical skills and no one that disbelieves it.
5. Analytics are employed and managed at the enterprise level.
6. They invent proprietary metrics for use in key business processes.
7. They become providers of analytics and data to distributors/customers and to suppliers.
8. "Test and learn" culture based on numerous small experiments on its DNA.
9. They commit to build internal analytics capabilities in the short and long run.
10. Analytics capabilities are turned into part of company's story and also into intangible asset shared in the annual report and in discussions with financial analysts.

Examples such as Thomas' 10 item list, and many other successful cases of implementing analytics, can be found here and in other articles on the subject.

This gives the reader a good precedent towards implementing analytics deeply and conversely. <u>Implementing Analytics</u> will always represent a mindset change for either an existing company or a startup who challenges the *status quo*. The effectiveness of such implementation will depend very much on the cost of the cultural change in respect to its benefits.

Let's take the World calendar as a metaphor for a second. Imagine the following situation; Some scientists propose a new day to start the World calendar. The new 1st of January would be the 21st of December as it represents the Solstice (winter solstice is in the north hemisphere and summer solstice in the south). There would be a small benefit as result of the change as it coincides with the start of the season: However, the benefit is just marginal compared to the huge cost of implementing it both from a technology perspective and from a historical perspective. The same goes with Analytics. The gain going forward (short and long term) of any change must always overcome the cost of changing it – nevertheless, as analytics is the biggest locomotive of change nowadays, when the world around us has already changed their calendar to start on the 21st of December the extra cost of not changing irreversibly moves to be in the hands of late adopters.

In the following chapters we intend to expand the vision of analytics beyond the views exploited in articles [1] and [2] and summarized in section 1 of this article. We intend to split Analytics Competitors into three types and analyze these three examples of companies who are competing in Analytics. The first types are companies with Analytics as core competitive advantage, (Analytics-Driven-Core) – this is the type many other articles like [1] and [2] talks about. We then intend to compare them with companies who were able to blend Analytics with their previous competitive advantage (Analytics-Driven-Blended) and finally we want to compare the previous two types of companies with Startups that have a serious commitment with Analytics (Analytics-Driven-Startup).

Chapter two is oriented to give beginners in the field some important background detail that will allow for the smooth reading of the rest of this article. Should you be an advance user of analytics, we recommend you to skip straight to chapter three.

2) LET'S GET DIRTY AND UNDERSTAND SOME OF THESE MODELS

Analytics is normally structured around explanatory, predictive and prescriptive models. In our context, a model is a formula or a set of algorithms that tries to imitate, explain and predict real world systems or behaviors. This chapter covers an example of a simple training structured dataset, and then it looks into four common model shapes. It touches on one possible model development process and ends by zooming in on a Credit Card operation to understand how a model is typically used for taking real time credit decisions.

<u>DATA SET:</u>

A typical structured model development dataset would contain a set of input variable and an outcome variable. The outcome variable represents the observation in reality of the phenomena under study. Typically the outcome variable is expensive to measure/capture or can have severe implications if one has to wait for it to be observable, this motivates the development of models to predict and/or to anticipate its occurrence. In the table below, the outcome variable indicates whether an individual has been diagnosed with a heart problem.

The input variables are typically characteristics correlated with the outcome variable, being much easier and cheaper to be collected.

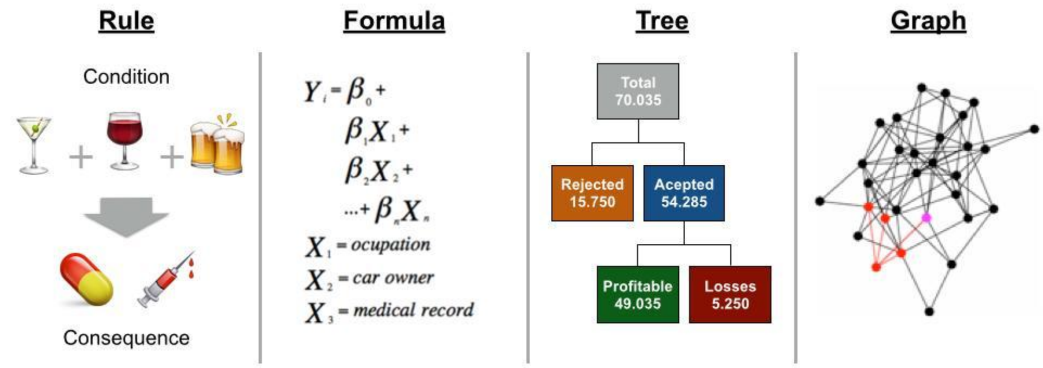
Input Variable 1 (Age)	Input Var. 2 (Profession Code)	Input Variable 3 (Weight)	Input Variable 4 (Hobby)	Input Var. 5 (Play sports?)	Input Var 6 (Blood Pressure)	Outcome Variable (Observed Heart Problem)
22	Student	100 Kg	Video game	No	140 systolic	1 (Yes)
67	Retired	75 Kg	Fishing	Walk	120 systolic	1 (Yes)
35	Lawyer	77 Kg	Tennis	Tennis	100 systolic	0 (No)

Therefore the model development process consists of generating a model which will associate a combination of input variables to define:

1. a probability of the outcome variable being 1 (YES) or 0 (NO)
   OR
2. a direct classification into group 1 (YES) or group 0 (NO).

MODEL SHAPE:

A model can be of many shapes depending on the technique and software used to develop it. The most common formats found are: rule, formula, tree and graph:



Rule models: there are many algorithms and models that use this format especially due to its power to explain complex systems in a very simple way. Market Basket Analysis (a priori algorithm) is one example of a rule-based algorithm widely used to train both unsupervised[3] and supervised[4] rules. A rule relates cause and effect or in rule terms a condition to a consequence.

$$\text{Condition} \rightarrow \text{Consequence}$$

A rule is said to be checked when its condition is true, what indicates then is that the consequence is likely to become reality. Many measures are used to assess the quality of a rule, the most common are: confidence, support and improvement.

Formula models: models made up of simple formula containing only sums, subtractions, multiplications and divisions have been in existence for over half century. It is widely accepted because of its simplicity and effectiveness. Logistic Regression is one example a formula-based algorithm which is widely used to train supervised models having a binary (0 and 1) outcome variable.

In mathematical terms a Logistic Regression model looks like this:

$$\text{SCORE} = \beta_0 + \beta_1{*}X_1 + \beta_2{*}X_2 + \ldots + \beta_n{*}X_n$$

Where the β's are called betas and are the model's parameters users need a software tool to help its estimation and $X_1 \ldots X_n$ are the observation of all input variables from a real dataset.

SCORE is the name sometimes given to the simple calculation of the formula above, while the probability of being 1 (YES) comes as result of the following exponential transformation:

$$\text{Prob(YES)} = 1 / (1 + \exp(-\text{SCORE}))$$

Tree models: probably the oldest model with its visual shape being a tree. The most common format of a tree found for modeling is a binary tree. Each node can have 0, 1 or a maximum of 2 offspring nodes. Tree algorithms, or Decision Tree algorithms as they normally called, are widely used to train supervised models having a binary 0 and 1 outcome variable and is especially popular when one is trying to identify outliers or using skewed data. One of the most famous algorithms to train Decision Trees is the C4.5 by Rossan Quinlan and can also be found under the name of J48 on its open source Java implementation.

[3] Unsupervised learning refers to a type of machine learning algorithm or statistical method used to draw inferences from datasets consisting of input data without labeled responses.
[4] Supervised learning refers to a type of machine learning algorithm or statistical method used to draw inferences from datasets consisting of a pair of input data and a labeled response for each instance of the data.

A Decision Tree training algorithm would have an expansion and a pruning phase based upon some predefined parameters typically a minimum support and a minimum confidence.

Graph models: graphs can be seen as a generalized version of a tree. These are formed of nodes and links where there is no restriction on link direction and number of offspring or even to number of parents. Any pair of node A and B can have one of the possible configurations:

- A B: no connection between A and B
- A → B: there is a connection from A into B
- A ← B: there is a connection from B into A
- A ↔ B: there is a connection from A into B and from B into A

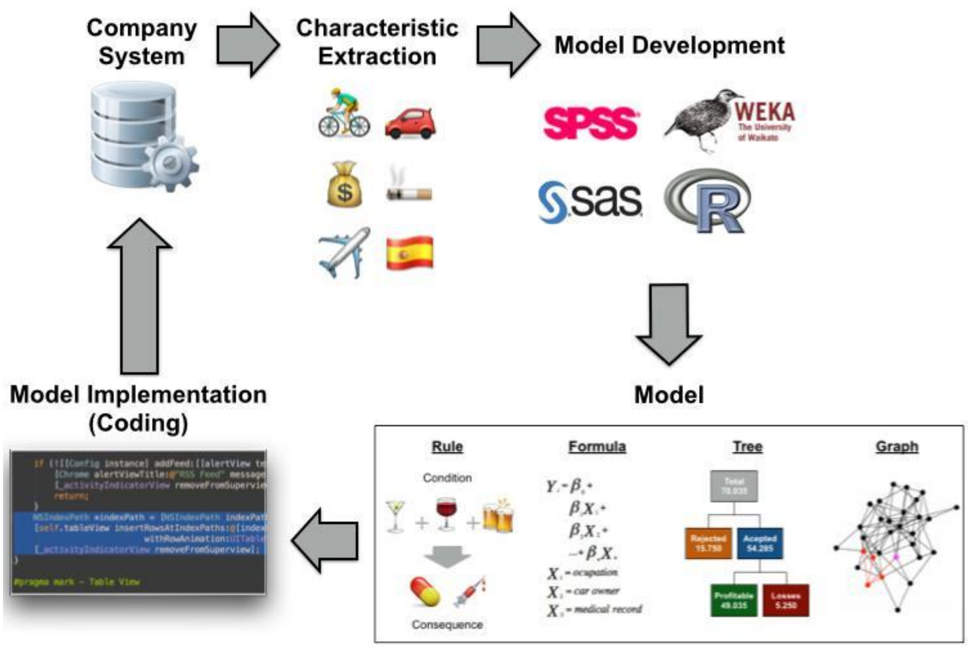
In comparison to other shapes, a typical graph training algorithm would require an extra input information describing connections between nodes (individuals or rows in a dataset). Examples of popular unsupervised graph models are HITS and Google's PageRank. These algorithms use information of the graph topology to output measurements for each individual node of its position, authority, and relevance.

MODEL DEVELOPMENT PROCESS[5]:

A typical model development process inside companies will involve:

- a human being
- a powerful PC or server
- and statistical software

Developing a Model



The process will start with the extraction and creation of the training dataset (this is the most time consuming aspect). Following this, many models are generated using the chosen model shape, technique and software tool.
The final step is where the human part is more important by choosing and agreeing the model that will then be coded into the company's systems.

Now that we have spent some time visiting four types of model shapes, a question that should arise in our heads is – Which method should I use?

[5] Please refer to http://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf for further reading on model development.

Rather than answering this question straight away, I would propose reflection. Experts with day-to-day challenges often forget to ask themselves this exact question.

There are some reasons to opt for one or another methodology, but in practice companies choose method 1 over method 2 not for a scientific reason but because of the folowing:

- It is a method vastly used by the industry they are in (conservatism).
- It comes implemented in the development software tool they have license to (amortized cost).
- It is the technique they are most familiar with and they know how to implement and monitor (change would incur learning curves and operational risks).

Whilst those are valid arguments, many times it forces a situation where the decision of the technique used for a model development to reside is not with the teams directly connected to the model development. Neither is it in the model use but it rests in the hands of many other stakeholders simultaneously (IT, HR, Finance, Senior Management), it creates an invisible and permanent stopper for evolution.

The same complex, invisible and permanent stopper for change happens regarding the choice of the statistical development tool. The list of available tools is very large – some of the more important examples are KnowledgeStudio, Model Builder, Oracle Data Mining, Revolution R Enterprise, Salford Predictive Modeling, SAS Enterprise Miner, SPSS Clementine, Stata, Statistica, TIBCO Spotfire Miner and WEKA. When analyzing the penetration of these tools by industry sector one can clearly see a dominance of some tools in some industry sectors. It is not uncommon to find historical reasons to have led a given tool to be adopted as industry standard of a given sector. Problems arise when this adoption becomes "Best Practice" and then it becomes the norm – stopping anyone from risking using a different tool.

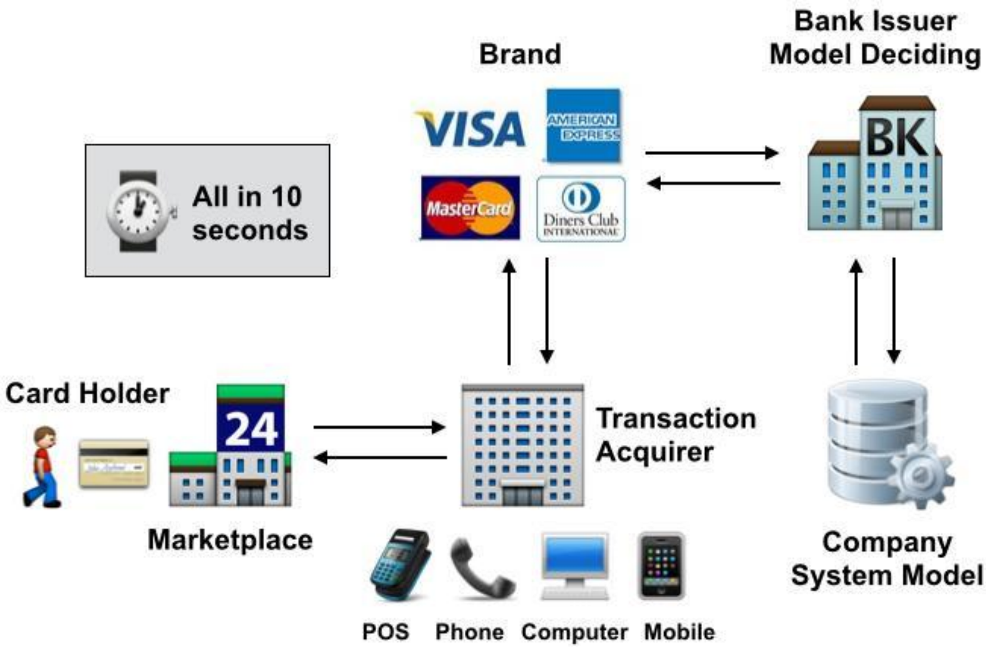
To answer the question in the mind of a beginner - Which method should I use? –
The only easy part of choosing a method is picking a supervised type of method. – a method that expects an observed value for the phenomena we are trying to predict, or an unsupervised type of method – a simpler method that has only input variables and does not expect an observed value for the phenomena we are trying to predict. However, going deep into the problem, most cases answering the question would require a costly, and therefore avoided, qualitative and quantitative comparison trying to find out the cost-benefit of using one technique or another. It is sad to realize that sometimes experts play a gullible role thinking techniques deliver similar results irrespectively of the application and the specificities of the data used at that moment (for example global crisis or financial booming). In rough terms, there is a set of techniques that tries to predict the average individual's behavior and therefore naturally outputs models that tend to be more robust (models that remain giving good predictions for longer). These models are more appropriate for situations/portfolio that does not change much – for example Bayesian Techniques, Regressions and Time Series. On the other hand, there is a set of other techniques that tries to separate different profiles. These techniques normally deliver models with better prediction, but those that lose their power quicker resulting in a shorter lifespan. These models are more appropriate for a skewed database or to search for outliers – Decision Tree, Neural Networks, Rule-based algorithms and Clustering.

MODEL IN USE:

Models are applied in many areas as we will see further in this article. With credit cards for instance, models are used to prevent the wrong use of a credit card by fraudsters.

Suppose a credit card holder approaches a store, the shop keeper swipes the card on POS machine (Point of Sale) and the POS machine establishes a connection with the Brand of the card (Visa, MasterCard, Amex, etc) this will then send the transaction to the bank that has issued the card.

Model at Work - Credit Cards



Inside the Issuer Bank the transaction goes through many checks like, card expiration, authenticity, limit availability and daily use limits. After this, our transaction finally reaches our fraud prevention model, at this moment raw data is transformed into calculated input variables and the model is then run delivering an output. Let's suppose in this output there is a very small probability of this transaction being fraud. So our model is prescribing the approval of the transaction.

If no other checks are carried out afterwards, the Issuing bank will respond 'approved' to the Brand, and the Brand will respond 'approved' to the original POS. All this happens within 10 seconds or less irrespective of where you are located on the planet and the location of your bank.

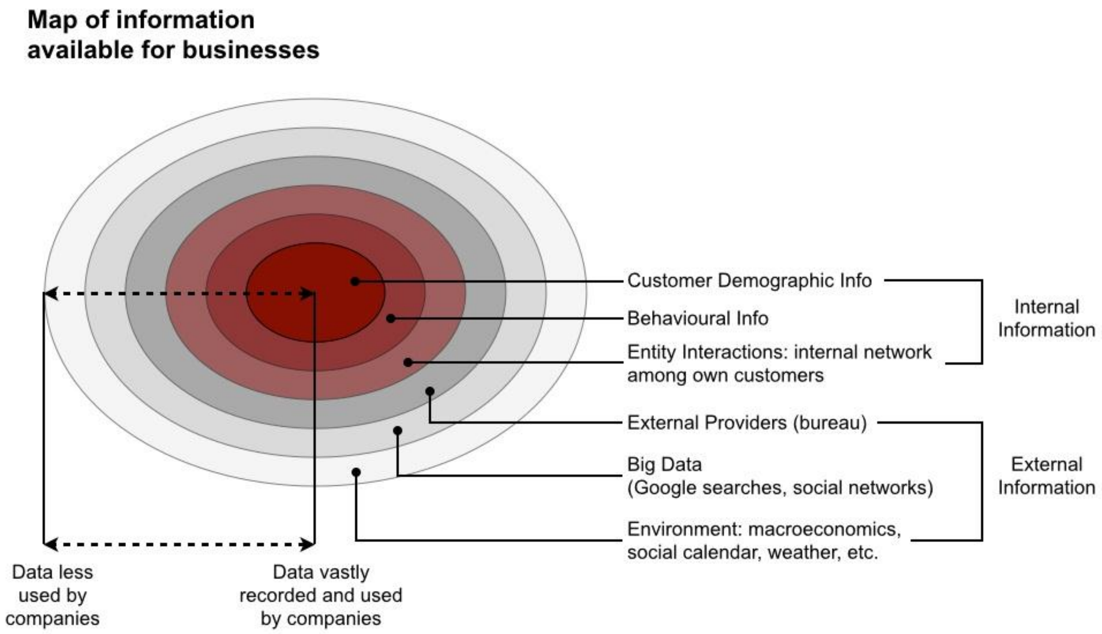
Two natural moves when modelers become more experienced with the problem they work on regularly is to introduce dynamic and prior elements into the current models. The introduction of the dynamic approach happens by: a) the use of time series, b) using methods that blend other techniques with time series (like turning logistic regression constant into a moving average and logistic regression parameters into dependence trees) or c) using tools that implement methods that allow business restrictions to be introduced mathematically (not as a posterior check) and that forces the method to output the best model making sure it makes total business sense (for example not predicting anything for under 18s or expecting that salary increases as age increases).

Finally, companies competing on Analytics do not limit themselves to the methods made available by statistical tools. They have a series of bespoke methods sometimes made by a consultant or self-engineered. The reason for this is simple, companies competing on Analytics are pushing the bounders with new techniques which means coding is necessary until the technique becomes available on a standard statistical tool. One recent example of this time lag is Social Network Analysis (SNA) methods. Google and other companies have been using Markov Chains and Google PageRank since the 1998 but it is only recently that statistical tools have released stable implementation of SNA techniques and will take many years for consultant companies to start selling these techniques broadly, let alone to businesses, taking its full potential on board. Similar to SNA, many companies have cutting-edge methods and bespoke models forming their competitive advantage and by the time the technique becomes public and available on standard tools it probably means they are using something else!

3) INFORMATION, MODELS MAP & SUCESS CASES

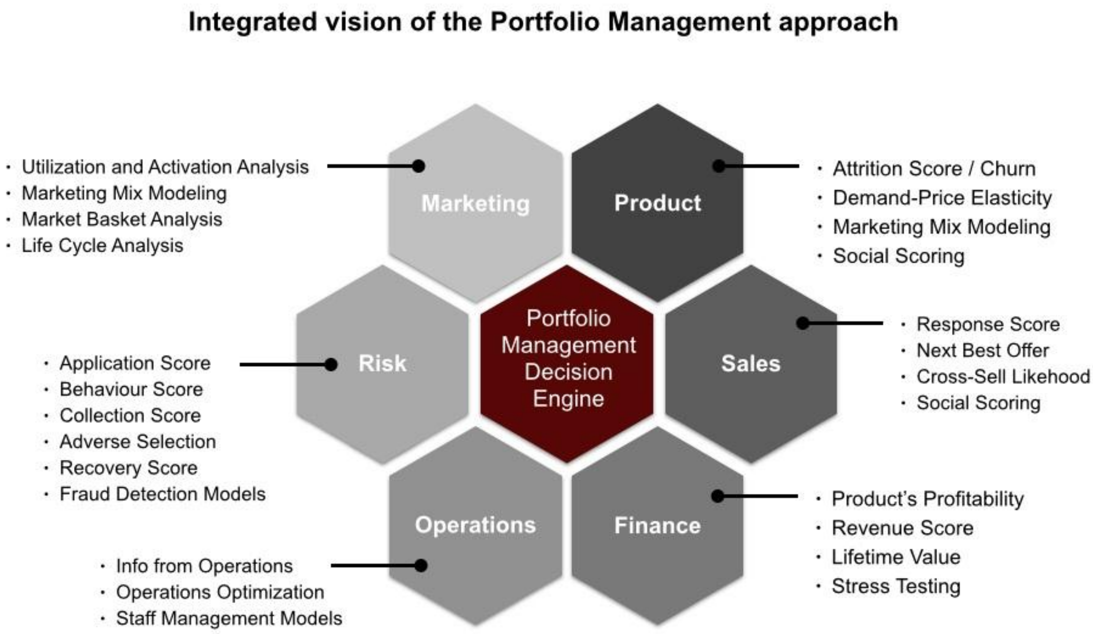
In this chapter we will review the type of information available for model development (and in most cases companies are not even aware that they could use it). We carry on discussing many types of models and their applications inside a company. We will conclude this chapter by displaying a table with many types of models that one single Analytics Driven company can be using simultaneously.

TYPE OF INFORMATION: The type of information used as input for these models were on a subject, that until recent years, was assumed to be very established and therefore not deserving much attention or investment from companies. However, reality is, Social Networks have being collecting valuable social data. Mobile companies have being recording valuable location data. Search engines have being recording people's preferences. The situation reached a tipping point when many providers started sharing or selling that information (normally anonymized). The new tendency became irreversible to a point where the new term Big Data was adopted to explain all this information available out there.



The diagram 'Map of use of information available for businesses' tries to summarize the information map available for model development. It should be read from the inner circle which represents company's internal information towards the outer circle which represents aggregated external information. Typically, companies start generating models using information from the inner circle and slowly move to the outer circle – this is the ideal movement - problems emerge because most companies jump phases. Phase three, the use of internal entity interaction to generate its customers Social Graph, is simply ignored as they jump straight into using External Data Providers. The next step that companies use after using external data providers is to try and integrate Macro and Micro Economics Statistics, once again companies jump the important phase of tapping into all available Big Data information.

MODEL VARIETY: Asking the right questions is key. Nevertheless, the right question will depend on many factors including: the industry being analyzed, the objective of the use of the model and the type of entity being analyzed. For the purpose of giving beginners an overview of the most common models a customer oriented business can have at each type of interaction with its customer, we have generated the illustration below (which is not intended to be comprehensive).



It becomes clear with this illustration that it is wrong to assume that Analytics Driven companies are "one trick ponies" that holds one single model that make them best in their fields. Reality is that companies like Amazon have much more than a very good recommendation system. Google has much more than an amazing sorting pages algorithm called PageRank and Capital One has much more than test-and-learn strategy.

Next, we detail few of the models from the 'Integrated vision of the Portfolio Management approach' picture:

- Risk Management Credit Scoring is applied to accept or reject new customers. They are done using a supervised method like a decision tree or a logistical regression. The input variables (or explanatory variables) will typically be the result of a form filled in by the new customer combined with additional variables the company will have internally or will acquire from information partners (like Credit/Information Bureau). The output variable will typically be profitable vs. non-profitable indicators;
- Market Mix Modelling is applied to give adjusted service to different groups of customers. They are typically done using a unsupervised method like clustering (K nearest neighbors for instance) but can also be used to predict specific responses . The input variables will typically be demographic information and, as it is unsupervised, no output variable is used;
- Social Scoring is very recent and is applied in many different fields. They are done using a blend of unsupervised, semi-supervised and supervised methods coming from Graph Theory. The input variables will typically be the connections among entities (calls, friends on Facebook, shared home/work address or shared phone number, and also transactional information like money transfer (P2P and P2B), bill payments and purchases at given point of sales.) The output variable will typically be increase vs. decrease in probability of the event under studied given the individual set of connections.

Below we change our viewpoint from the 360º angle of the single company we discussed in the previous section and present a table with an overview of prominent models from companies from various sectors.

SUCCESS CASES		
FUNCTION	**DESCRIPTION**	**EXAMPLARS**
Cross Selling	Use recommendation algorithms to mass-customize the online store for each customer by posting the most relevant offer to each customer	**Amazon** (USA & Worldwide - Amazon Recommendation System)
Customer selection, loyalty, and service	mycokerewards.com - Understand end customer needs; increase product adoption; retain their loyalty	**Coca Cola** (USA & Worldwide)
Demand prediction	Vivo & National Express both decides capacity allocation, pricing and monthly revenue using demand forecast.	**Vivo** (Brazil mobile phone); **National Express Group** (UK's leading transport provider)
Financial Performance	Understand the short, medium and lifetime drivers of financial performance and the effects of nonfinancial factors in the company's bottom line	**Barclays Card** (UK), **Capital One** (USA), **Credicard** (Brazil), **Thinkoff CSB** (Russia)
Human Resource	Select the best employees for particular tasks or jobs and define best compensation packages to attract talent.	**Oakland A's** (USA) **[4]**
Improve Service	Alter the relevance and the display order of pages resulting from web searches and also offer Analytics to users especially for ad users.	**Google** (USA & Worldwide - Google Pagerank)
Pricing	Identify price elasticity that will maximize revenue and minimize inventory, especially during price reduction campaigns.	**El Corte Ingles** (Spain - biggest department store group in Europe)

4) ANALYTICS-DRIVEN-CORE VS. ANALYTICS-DRIVEN-BLENDED VS. ANALYTICS-DRIVEN-STARTUP

When undertaking the implementation of Analytics each company will almost certainly experience their own challenges, in addition, their motivation towards implementing Analytics also varies hugely. In this chapter we introduce a new angle for looking into Analytics-Driven companies. We will split the companies that have successfully implemented Analytics as competitive advantages into two groups and add to that a third group of Startups. The aim of this categorization is to unveil how different the road map for each group would look. Such differences could even drive some people to doubt that they are implementing the same thing.

The first type are companies with Analytics as core competitive advantage, (we call them Analytics-Driven-Core). These are the ones many articles similar to [1] and [2] discuss. We then compare those companies to companies who were able to blend Analytics with their previous existing competitive advantage, (we call them Analytics-Driven-Blended), and finally we want to compare the two previous types of companies with Startups with a serious commitment with Analytics (we call them Analytics-Driven-Startup).

a) Describing each type

There are companies out there who really have mastered the art of number crunching in their fields. Nowadays, we see examples of Analytics applied to a vast number of fields, that in the past were just a privilege of banks, telecoms and retailers. Many years ago one could attribute Capital One's success to them being a mono line Credit Card issuer, or being the first mover with the test-and-learn strategy or for simply being lucky. These days, not only are there dozens of literary examples that confirm they followed a winning track, but there are many examples around the globe of companies including the UK based Barclay's Card, the Brazilian based Credicard and the Russian based Thinkoff Credit System Bank that successfully implemented strategies similar to Capital One and reached success of similar or larger magnitudes on their local markets. Clearly all Analytics-Driven companies have Analytics in their DNA, and from top management to new starters all staff are hand-picked because they take decisions on fact based style, but as we will see next we are pushing too much to allocate them all into the same bucket.

Analytics-Driven-Core

What really differentiates an Analytics-Driven-Core company is the role Analytics plays. Being the bedrock of the company it is a fundamental part of the company's competitive advantage from early days. Hundreds of statistical and mathematically developed models work side by side with thousands of tests. This enables this type of company, to translate into atomized procedures all expertise and experience of their staff. Assessing every aspect of their business, in a constant search and optimization process, they seek to transform themselves into their best version for all stakeholders. (Examples being: customers, employees, shareholders, government, planet).

An easy example to understand an Analytics-Driven-Core is probably Google. If one removes Analytics from Google - the models they use, like PageRank to sort pages before displaying it – even though, the great amount of information and powerful servers would remain, Google service would very likely become worse than competitors driving Google to lose its market leadership. Therefore, Analytics is at the core of its competitive advantage.

For example: For Credicard in Brazil, Analytics enabled the company to vanish all manual credit underwriting decisions (zero grey zone / zero overrides), and also made it possible for them to create a virtual business (no brick-and-mortar business) before the Internet existed by eliminating all bank branches and making call centers the only channel customer could interact with the Company. In the year 2000, Credicard celebrated its 30^{th} anniversary and its many years as the market leader for credit cards in Brazil, holding more than 30% market share. The company was so information oriented that it invested millions collecting and storing information. They eventually reached the point of holding information on more than 110 million Brazilians, representing more than 90% of the entire Brazilian work force at the time. Holding information on each individual, irrespectively if they were an existing Credicard customer or not, they were able to predict with models various things such as credit worthiness, offer take up probability, average credit card use (in $), acquisition cost, best channel to contact and revenue estimations.

All these together created the input for the most important model of all. Credicard were able to calculate for all existing customers in their database, and all potential customers in its mailing, a price tag or in their jargon the customer Lifetime Value (LTV) using more than 20 years of year-to-year projections. Credicard Lifetime Value for each customer was so precise that during a company valuation exercise in 2005 the valuation using standard financial methods and the sum of all customers LTV ended up being less than a 5% difference, both being around 4.5 billion Brazilian Reals (around 1.67 billion US$ at the time).

As one may imagine, Analytics-Driven-Core companies are typically very fact based and structured on their decisions. This can be observed in the decisions a company takes in relation to services or facilities that they offer or withdraw from their customers, in meetings of managers or even in super-integrated systems. They are growth orientated and are controlled risk takers.

Analytics-Driven-Blended

As Analytics travelled across to other fields, established companies with existing successful competitive advantage were able to blend Analytics with it, thus creating even more powerful corporations turning them into Analytics-Driven-Blended. It is not difficult to imagine that it was done following a completely different path from the previous group.

Coca Cola company is a great example of how an established company with many competitive advantages from brand awareness, size, penetration, proprietary formulas, to name a few, can have very little information about their final consumer. Using the platform mycokerewards.com, coke was able to turn it around and became so knowledgeable about the end consumer that they start selling this profiling down to their distributors and retailers making the Analytics a profitable business unit in itself.

Carrying on with Coca-Cola, it is an great example to understand the Analytics-Driven-Blended concept. If one removes Analytics from Coca-Cola - the models they use, for example Market Mix Modelling and demand prediction – the company would very likely lose an incredible amount of sales, some locations may start having problem receiving the fizzy drink and costs may skyrocket, but Coca-Cola would remain being the only one selling Coke and competitors would still have a hard time taking its leadership away.

Analytics-Driven-Blended are very fact based with a more long term result approach and most of the time with a very strong cost-driven orientation. They tend to use Analytics first to enhance the reach of their existing competitive advantage by making it more replicable and by reducing it cost. Another strong use of Analytics is for defensive strategies especially on price and discount strategies. These types of companies tend to have more investment in Analytics for customer retention in comparison to Analytics-Driven-Core and Analytics-Driven-Startups. They are orientated in defending market share and increasing profitability, they tend to be risk avoiders.

Analytics-Driven-Startup

A recent big wave of Analytics driven competitors is coming from the field of technology startups. Analytics-Driven-Startups have a clear analytical mind set merging passion for data crunching with pivoting flexibility. In many case they try to repeat the success of other tech companies Analytics success for instance Google's Pagerank and Amazon Recommendation System.

A recent example is a company called Flurry. Flurry offers a piece of code (for free) to Mobile App developers so that they can measure the size of the mobile app market they are on and compare the number of downloads their App has to their peers. Everything seemed just fun inside Flurry office until the pivot moment.

The 1 billion mobile devices using Apps that uses Flurry code and a powerful Analytics team enabled them to develop an understanding of the profile of mobile users that no one else has. In 2013 they unveiled their real money making formula: Flurry Marketplace, a real-time bidding exchange for ads in mobile apps. Demand-Side Platforms (DSPs) and Agency trading Desks (ATDs) now they could programmatically reach their intended audiences without having to rely on third party audience data. With a single point of integration, advertisers could connect to Flurry Marketplace to reach more than 300 million monthly unique users and bid on billions of ad requests per month. How good is that for a Startup?

Analytics-Driven-Startups, should they be a recent created company, one that has acquired some growth over the years or a spin off expending their perimeter of customers, they all are seeking actively their repeatable and scalable business model. This seeking puts them in the position of a risk taker and the majority of the time puts them in a very aggressive position to grow on their niche market. Analytics-Driven-Startups have to be the most agile of them all. Often they are the first movers to enter a market niche – frequently they have to be testing many entry strategies in several different niches at once, they also have to be alert to decide when to become hostile towards competition as this sort of confrontation is a startup business as usual (BAU). All of this forces startups to be the most vigilant, requiring therefore a different set of analytical skills and capabilities. Rather than long standing statistical models or a structured set of key-performance-index (KPI), startups make their bets on quick-to-implement statistical models that give extra performance. Most importantly, they invest on report tools that deliver a low cost, dynamic and very agile ad-hoc reports as it is not rare the situation that where these analysis will be used only once and never again. Being the quickest to enter a market niche is not a complete advantage, if the startup is not vigilant and the quickest to expand investment, or more so, the first to exit that market if necessary. This entering-exiting paradigm is the key element for a successful startup and for that reason receives a special name **pivoting**.

In order to discuss important issues for startups, next we set the scene for Lean Startups and reserve a sub-session to discuss the problem of no-data to start with. We will discuss the example of 'banco ibi' that evidences the help Analytics can give for agile pivoting and many examples of companies overcoming the no-data issue.

b) **Lean Startup: enemy or ally of Analytics?**

According to Steve Gary Blank, “a startup is a temporary organization formed to search for a sustainable, repeatable and scalable business model”. These companies, generally newly created, are in a phase of development and research for markets. The term became popular internationally during the dot-com bubble when a great number of dot-com companies were founded.

In this session we will center our attention on the section saying: “search for a … repeatable and scalable business model”.

Steven Gary Blank, in his book, The Four Steps to Epiphany **[3]** summarizes on four pillars, the pattern he found on startups that succeed. 1 - Customer Discovery; 2 - Customer Validation; 3 - Customer Creation; and 4 - Company Building. Later, one of Blank’s student named Eric Ries, put together a method called Lean Startup which aims to provide startups with a method to search for a repeatable and scalable business model by mimicking what previous successful startups have done. Mr. Ries’ method is divided into 5 pillars: 1 - Minimum viable product; 2 - Continuous deployment; 3 - Split testing; 4 - Actionable metrics; and 5 – Pivot. <u>This is where Analytics connects with the combined work of Mr. Blank and Mr. Ries</u>.

For Startups that are committed to Analytics, Analytics can be the fuel for Lean Startup, in other words, the main tool for the search of a repeatable and scalable business model. However, for Statups not selling Analytics, having first class Analytics does not a guarantee this success. Analytics cannot be the main element of a Startup competitive advantage but the main driver for pivoting. One interesting example that evidences this is banco ibi in Brazil. Banco ibi was a spinoff of the European clothing retailer C&A. ibi originally was a department inside C&A stores managing all private label credit cards. With the spin off they started to offer financial products, among them, credit cards, loans and insurance not only to C&A customer as before but to everyone in Brazil. At a first glance the change seemed logical, but it did not take long for them to realize that they had no advantage compared to other companies in the open market and in many channels their marketing and sales cost were higher than operational income. However, the company had a great secret weapon, instead of many standard reports or long-lasting statistical models or an error free IT governance, they had a very agile data cube, a very agile reporting tool, an analyst who knew how to program and a very agile IT implementation process. All these combined gave them the agility they need to run dozens of brand new ad-hoc analysis every week, roll out new models within days, implement hundreds of new tests every month, and more important, remain serving profitable niche markets and move out from those making losses. They were so creative on their “finding a space” saga that after around six years of existence, the company became known for its debt collection and recovery efficiency and started profiting hugely from it by buying bad debt portfolios from competitors. For Banco ibi Analytics were just the natural way of pivoting, without even realizing they were pivoting, in a search for their repeatable and scalable business model.

c) **Startup Analytics - An oxymoron?**

A handicap that Startups face when competing with established companies nowadays is the fact they have to start having zero data to set up strategy.
In spite of this we see more and more cases of Analytics-driven-startups that surprise us with creative ways of overcoming this problem and becoming very powerful competitors.

One clever way Startups have found to collect information and quickly enrich their databases is by designing an entire business model to do this task. An example of this is Lenddo. Lenddo is a startup micro-finance firm set up in New York but oriented to the emerging markets with business in Philippines, Mexico and Colombia. Lenddo claims to provide loans to the wide middle classes using micro-finance lending techniques of the last century but with a technological and social twist. It holds a membercentric credit reputation scoring system leveraging data from social networks like Twitter and Facebook. They also established a vouching system for online contacts to support the applicant requesting the loan. Clearly, Lenddo’s acquisition process forces applicants to share their basic social network information with the company and with this alone Lenddo has enriched their database to assess the applicant. They can potentially gain a basic credit assessment of the applicants' social connections as well. By requesting the applicants connections to vouch for them, Lenddo is forcing, first month-to-month publicity and is also trying to further enrich their database by acquiring social information of the people vouching for the original applicant. Even if Lenddo finally rejects the applicant it will have collected valuable information that can be used in many more credit assessments.

Another clever way Startups gather information and quickly enrich their databases is by designing a parallel business orientated to enrich data which is then used by the real business. We come back to the example of Flurry from section 4a. Flurry offers, for free, a piece of code that is very useful for Mobile App developers. It collects information of users downloads of their apps. On the other hand, for Flurry, this is the way that they collect information for their real business which is to offer a platform for a real-time bidding exchange for ads in mobile apps.

For a final example, we use again the data enrichment from Coca-Cola. As an established company with many competitive advantages, but with close to zero information about their final consumer, Coke was able to turn it around and become so knowledgeable about their end consumer that they started selling this understanding down to their distributors and retailers making the Analytics a profitable business unit on its own.
Using the platform mycokerewards.com, my coke reward case shows how a factory that sells to distributors and has no contact with the end customer cleverly creates a system to collect end-customer information and finally start selling Analytics to their network of distributors and retailers.

5) IRONY OF LIFE – THE MODEL RISK MANAGEMENT

Financial services were pioneers in applying Analytics to master their fields, but we first saw a change in focus from predictive power for more robust models arising from within the organizations. More recently, we are seeing a movement towards a more robust management of the entire lifetime of a model pushed by financial regulators. Ironically, less and less the competitive advantage different modeling approaches can give in this industry, is losing site for a more standardized approach in order to make the banking system more robust. A clear example of this is the focus giving to Model Risk Management. "Model risk" is the risk that the predictions may be incorrect, causing a company to incur unexpected losses. Consequently, model risk is a broad topic covering many elements of risk management and providing obvious challenges regarding consistency, robustness and quantification of the assessment process. Model risk arises from: the limitations of the development data; erroneous modeling assumptions; incorrect use of tools, techniques and methods for model development; delayed implementation of the model; incorrect use and operation of the models; consistent and material manual overrides of the models outputs; inadequate monitoring, validation and update of the models. According to the American Federal Reserve, *"Model risk management begins with robust model development, implementation and use. Another essential element is a sound validation process. A third element is governance, which sets an effective framework with defined roles and responsibilities for clear communication of model limitations and assumptions, as well as the authority to restrict model usage." (The Board of the Federal Reserve, Ref 1, p5, para 3)*

It is clear that companies must comply with regulation and we support that. Nevertheless, when a company commits to being Analytics-Driven it is key to keep reminding itself of its priorities. A company seeking to be the best analytical company in its field cannot base their success uniquely on industry standard software solutions, even if it pleases Regulatory Authorities, because it clearly can cost its leadership.

6) REMARKS AND CONCLUSIONS

In this article we have discussed how companies blended Analytics as part of their competitive advantage to become much stronger players while including lots of real world examples. The real examples used leave a clear footprint showing that Analytics add a great value to companies where it is applied, especially in the long run and in achieving and defending leadership. Rather than being a success case study or a text book for Analytics, we aimed this article to look into the road map that takes a company to be a successful Analytics-Driven competitor and we could not find one but at least three different paths. Therefore, we included a new angle for understanding companies that have mastered Analytics by separating them into three groups: companies with Analytics as core competitive advantage; companies who were able to blend Analytics with their previous competitive advantage and look inside Startups that are committed to Analytics. We reserved space and discussed why, for the majority of Startups, Analytics cannot be the sole competitive advantage but the main driver for pivoting.

By analyzing all the previous aspects of Analytics in real companies, we can only conclude that being Analytics-Driven is a way of life. Becoming an Analytics-Driven company for the cases analyzed were an exercise of willingness, commitment and investment. Independently of size or stage of the company, they decided to marry Analytics by applying it to all parts of their business and most importantly it was a long term commitment.

For companies willing to start this journey, we are sure that the path will be full of temptations trying to drive them away from the correct road map. Therefore, keeping remind themselves of this commitment or even getting to the point of tattooing Analytics in the heart of the company's culture values would ease the way. It is a fact that the believe on Fact-based decisioning, modeling and test and learn must be shared by all people in the organization and if we have to choose one group of people that should champion this more anyone else they would be the Senior Management. It should be Senior Management priority to have this believe in the heads and hearts of everyone and to only select people with the very best analytics skills. Each individual must feel that each decision they take big or small must be data-driven - every day – no exceptions. If a company reaches maturity in Analytics, this faith in Analytics will naturally transcend their walls and be shared with customer, distributors and providers either by influencing/helping them on the same path or by associating only with entities with shared values.

Finally, linking to the title of the article, you should fear Analytics driven competitors because they are simply becoming leaders in every field nowadays. The biggest sign of this is the Analytics ‘acqui-hire’ that is taking place in many industries. To name just few of these acquisition/hiring: Accenture acquired Neo Metrics, Facebook acquired Atlas ad Analytics, Google acquired DeepMind Technologies, IBM acquired Star Analytics, Walmart acquired Inkiru, BBVA acquired Simple Bank and the list goes on and on. Price tags of these deals are astonishing, for instance BBVA paid US$117 million for Simple Bank and Google paid US$650 million for DeepMind Technologies. These huge valuations are never justifiable only by looking into traditional valuation criteria and taking into account the value of technology, data, customers and potential growth. Prices are showing to the world a new trend, a trend where immense premium are being paid for the brains and the culture of these acquired companies and this can only persist because in most cases the acquired senior manager team is signing contracts to remain and continue leading the new company as semi-independent businesses.

References:
[1] Case Study: Competing on Analytics, Thomas H. Davenport, Harvard Business Review , R0601H-PDF-ENG, Jan 1, 2006
[2] Capital One Financial Corporation: Setting and Shaping Strategy, Garth Saloner and Victoria Chang, Stanford Graduate School of Business, dic 17, 2004
[3] The Four Steps to the Epiphany: Successful Strategies for Startups That Win, Steven G. Blank, Edition 3, Publisher Cafepress.com, 2007
[4] Book: "Moneyball: The Art of Winning an Unfair Game", Michael M. Lewis, Edition reprint, Publisher W.W. Norton, 2003