BarryAnalytics Limited

Embedding Advanced Analytics into Business Applications

Dr Barry Leventhal
and Dr Stephen Langdell

www.barryanalytics.com
Transforming Data

Introduction

There can be little doubt that the use of Advanced Analytics is increasing at a rapid rate – industry researchers all agree that Analytics is a growth market. For example, according to Gartner[1], predictive analytics software will be pervasive by 2020. By *pervasive*, they mean that three-quarters of all users of Business Intelligence (BI) systems will have access to predictive functionality. Today, that figure is less than one-third; Gartner expects it to rise to half of all users in the next two years.

The reasons for growth are clear. First and foremost, the volume of data being generated is increasing at an exponential rate – companies now want to store, understand and capitalise on this information. According to Forrester Research, the customer data explosion shows no signs of abating[2]; "*Companies need strong customer analytics capabilities to get a handle on customer data and make meaningful decisions.*" Other drivers of growth are the fact that computing power has become more affordable, greater importance is being placed on analysing the value of each customer, and more software products are available to aid this process.

Decisions based on Analytics need to be taken more quickly and seamlessly. In a recent book[3], Davenport and colleagues show how companies apply Analytics in their daily operation. They discuss what to consider when introducing Analytics into business processes, and point out that the effects can be profound. Similarly, a research brief[4] on embedding BI in enterprise applications by the Aberdeen Group suggests that this results in enhanced business performance.

The theme of our article is to show how advanced analytics techniques – such as forecasting, predictive modelling and optimization – may be embedded into business applications. When integrated within an application, such as a recommendation engine that selects the next best offer to make to each customer, the results can often lead to improved recommendations and higher return on investment (ROI).

As we will explain below, companies are now able to create their own business tools and applications, containing embedded advanced analytics drawn from a large library of mathematical and statistical algorithms.

Challenging problems require more integrated analytics

"Yesterday's approach" – which most users follow when employing analytical models – is to build and deploy each model as separate stages. There's nothing wrong with yesterday's approach – it's the method advocated by the CRISP data mining process[5], and adopted by many providers of analytics and data warehouse solutions. In our experience, it's the best way to develop models initially, in order to understand and demonstrate their benefits.

The main software options for "yesterday" include statistics packages, data mining solutions that can handle larger volumes of data and more automated model development and deployment, and statistical programming languages.

However, today's challenging business problems require a greater degree of integration between the model and the business process, and automation in both model development and deployment. Why do we say this? Because the data to be modelled often changes rapidly – for example, a retailer introduces new products on a weekly or even daily basis. So if a separate model is required for each product in order to optimise its selling price or cross-sell to customers, then both the model development and deployment need to be handled automatically within a business application as part of the operational process.

Table 1 provides examples of business applications that are likely to require or benefit from embedded advanced analytics, together with some of the techniques that would typically be involved.

Table 1: Example business applications of embedded advanced analytics

Business Application	Question	Techniques
Price optimization in Retail	How should products be priced in order to maximise overall profitability?	Price elasticity models, Optimization methods
Markdown optimization in Retail	How should perishable items be marked down, in order to minimise wasteage?	Price elasticity models, Optimization methods
Portfolio optimization in Financial Services	How should an investment portfolio be constructed and reviewed?	Optimization methods
Credit scoring	Which applicants/customers for credit are likely to default?	Logistic regression, discriminant analysis, decision trees, neural network, support vector machine
Supply/demand chain forecasting in Retail	How to plan the movement of items through the supply chain to maximise availability and minimise inventory levels?	Forecasting models- simple linear regression, regression with smoothing, ARIMA models
Capacity planning, e.g. for call centre management	How to forecast demand and allocate existing resources?	Regression models, neural networks, Optimization methods
Customer relationship management (CRM)	Which product/service is each customer likely to purchase next?	Autoregressive models, maximum likelihood estimation
Propensity modelling in CRM	Which customers are most likely to respond to a marketing campaign?	Decision trees, logistic regression, neural networks, support vector machine
Customer contact management	How should customers be allocated to a set of marketing campaigns, in order to achieve objectives and satisfy constraints?	Optimization methods
Customer churn/attrition management	Which existing customers are most likely to churn/attrite?	Survival analysis models, autoregressive models, maximum likelihood estimation
Lifetime value/duration management	What's the lifetime value of each customer? How long before each customer becomes likely to churn/attrite?	Survival analysis models,proportional hazards models
Marketing mix in Retail	How to maximise the income from a product by combining product specification, distribution channel and promotional tactics?	Quadratic optimization with linear constraints
Loan offers optimization	How to optimize and validate financial loan terms?	Constrained optimization
Product lifecycle planning and forecasting	How to plan a product lifecycle scenario to plan investment returns and avoid risky financial decisions?	Solvers for systems of nonlinear equations, probability distribution functions

Approaching today's challenge using an algorithm library

Today's and tomorrow's approach is to embed the mathematical and statistical techniques in a business application. This can be done in three different ways, according to the theme *build*, *borrow*, or *buy*[6]. Firstly, the user can *build* a software solution from scratch. Although the functionality is fully owned by the user, it requires time and effort to write, test, document and then maintain the underlying code-base. Alternatively, the user may decide to *borrow* a solution, which means embedding freely available open-source analytics. Although this saves development time these can vary in their quality, documentation and support. The third approach is to embed commercial software, such as the NAG Library[7].

The Numerical Algorithms Group is a software vendor, whose numerical library has been well-known for over forty years. With over 1600 methods[6] it is the largest commercially available numerical library that spans areas of numerical analysis and statistics, such as:

- Optimization
- Solvers of linear/nonlinear systems of equations
- Interpolation & Approximation
- Correlation & Regression Analysis
- Random Number Generators
- Nonparametric statistics
- Survival Analysis
- Time Series Analysis

The library comes together with thorough and extensive documentation and an excellent support service. In addition, NAG helps businesses with services and consulting to deliver high quality bespoke solutions.

Case study example – propensity modelling for customer management

Companies with large customer bases often use propensity models, for example, to target products/services to the best prospects and also to identify which customers are most likely to leave them – so that appropriate retention actions can be taken.

A variety of alternative modelling techniques may be employed when developing a propensity model including Decision Trees, Logistic Regression and Neural Networks. One primary consideration, when choosing which technique to use, should be to understand how accurately each method will predict the outcome to be targeted. This requires building and evaluating a set of propensity models using all of the available techniques – which can take up a large amount of analyst effort, and so is rarely undertaken in practice.

Model comparison application

In order to show how this question may be answered using embedded model algorithms, BarryAnalytics and NAG created a model comparison application which given a dataset, will fit alternative models using different techniques and then compares their performance. The application splits the dataset into development and validation sub-samples, and then fits models using four techniques available in the NAG library:

- Decision Tree
- Logistic Regression
- Neural Network (Radial Basis Function)
- Support Vector Machine (SVM)

Decision Tree and Logistic Regression are straightforward methods with relatively few parameters to tune; these techniques are widely available and frequently used by marketing analysts. Neural Network is a more sophisticated approach which improves on the standard methods in situations where a non-linear model is needed. Support Vector Machine is a relatively recently developed advanced technique, which is designed to handle highly non-linear relationships.

Each technique requires a small number of parameters to be tuned, in order to control its operation – more so, for the Support Vector Machine and the Neural Network. The software solution employs sensible default parameter settings, although any of these can be over-ridden to observe the impact that they make.

Model performance is displayed using gain and lift charts, based on the validation sub-sample – examples are shown in Figures 1 and 2. The tabular results for both the development and validation sub-samples may be output to a spreadsheet for more detailed interpretation.

Figure1: Example gain chart

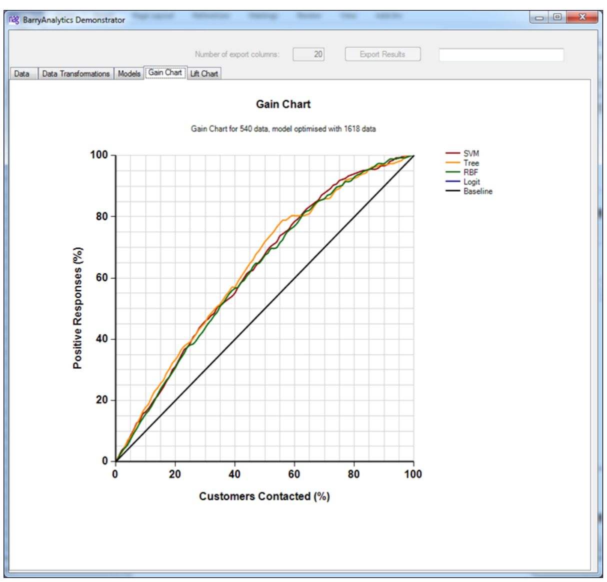
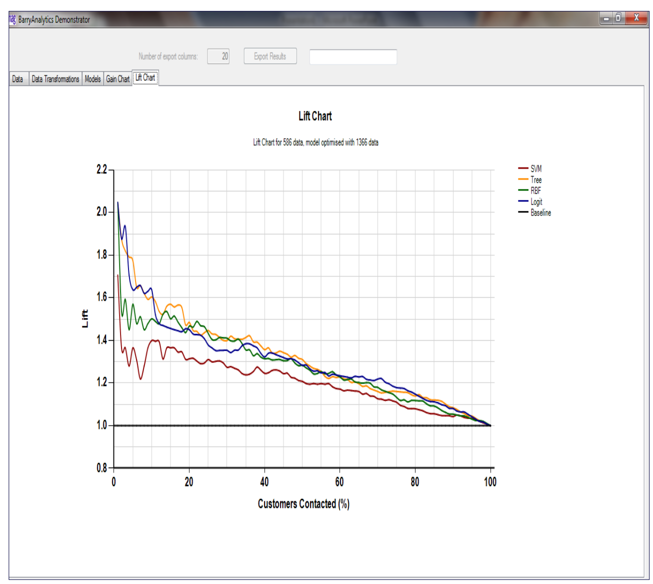


Figure 2: Example lift chart



The application was developed using Microsoft Visual Studio in the .NET framework linking to the NAG Library. The library can be called from any of the .NET languages (C#, F#, and VB.NET) as well as from other programming languages and environments such as C/C++, JAVA, Python, or Excel. The programming language used in this application is C# and the graphical user interface is a part of Visual Studio suite. This combination allows for rapid prototyping of applications.

Example use case: Risk-based verification in the public sector

Risk-based verification is an approach used by local authorities to assess claims for housing and council tax benefits. It enables a local authority to assess the likelihood that a new claim contains potential fraud or error, by calculating a risk score – the score indicates the level of checking that needs to be applied to that case.

An initial risk model had been developed for a group of local authorities – the model distinguished well between low, medium and high risk claims.

Our model comparison tool was applied to the analytical dataset used for the risk model development, in order to examine which analytical technique might result in the most powerful model. A number of runs were carried out, in order to find suitable parameter settings for each technique on this set of data.

The results suggested that the Logistic Regression and Neural Network would be the most powerful techniques, out of the four algorithms examined, in order to target within the highest 35% of risk scores. Decision Tree came next, and gave similar discriminatory power to the first two techniques if targeting across more than 35% of risk scores. The Support Vector Machine showed lower levels of discrimination than the other techniques at all levels of targeting.

The results are illustrated in Figure 3, which compares the indexed lift ratios of the four techniques for the highest 20% of risk scores.

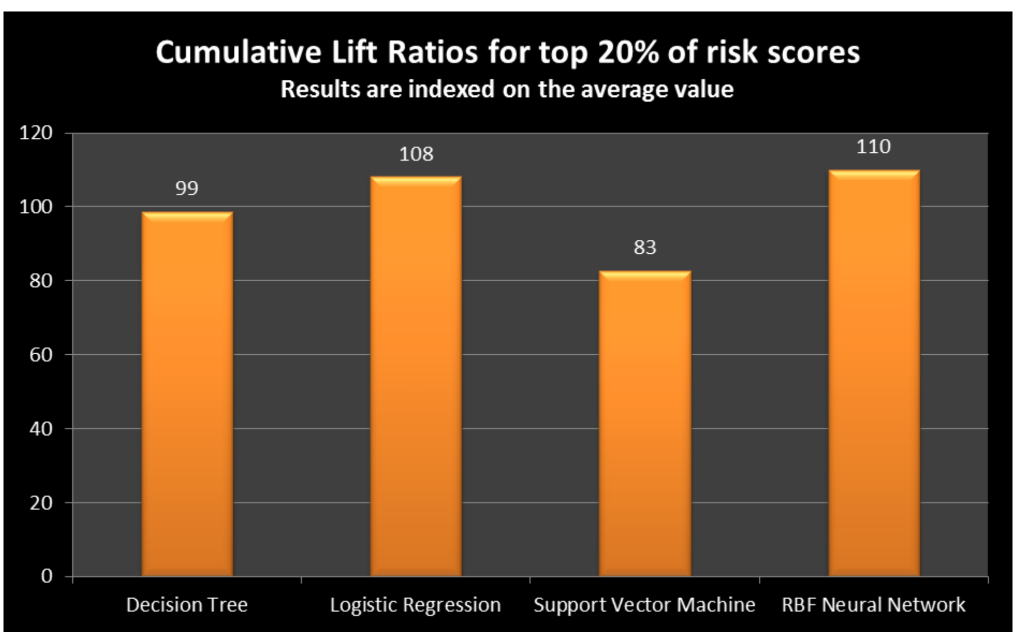
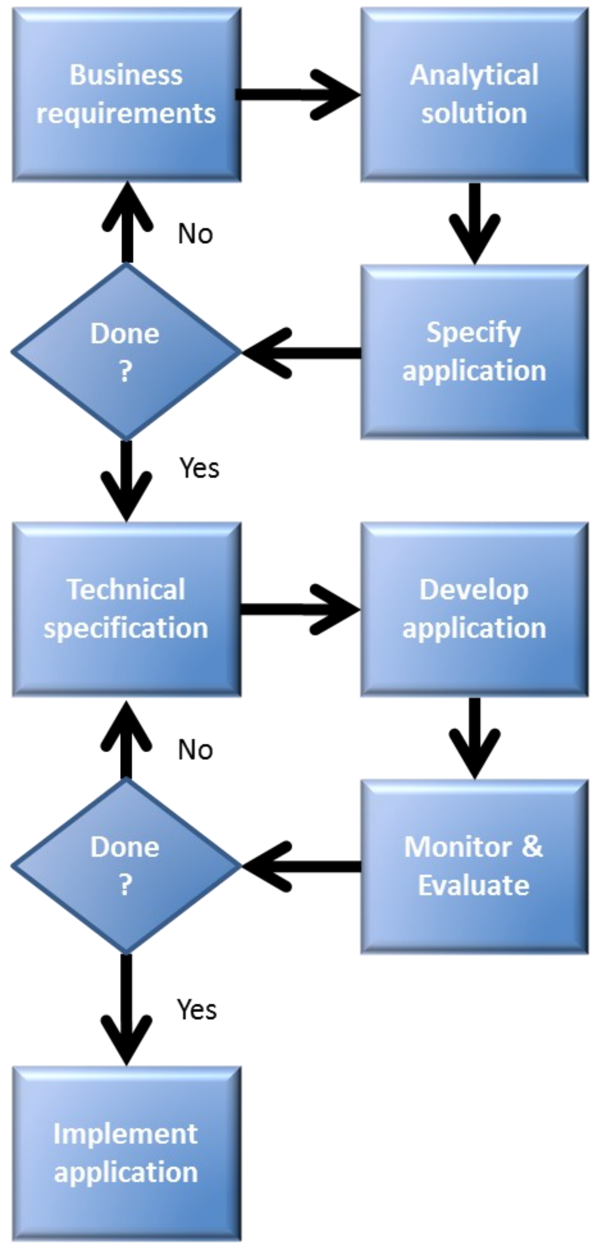


Figure 3: Example results—comparison of models for risk-based verification

Based on these results, one would recommend that the modelling techniques to consider for this risk-based verification problem would be either the Neural Network or Logistic Regression. For model transparency and ease of implementation, the regression approach was selected.

Figure 4: Application Development Pathway



Developing an application containing embedded analytics

Three skill sets are usually required in order to develop an application that contains embedded advanced analytics algorithms:

- A detailed understanding of the business problem, the requirements for a successful solution and how the solution will be applied. This knowledge comes from the business user.
- The ability to create the analytical solution, if this does not already exist, and to translate the solution into a specification for the application. These skills are supplied by BarryAnalytics (if they are not available in-house).
- Software development skills to develop and deploy the application. These skills are provided by the consultants at NAG (if not available in-house).

Figure 4 summarises the typical stages in building a business application that embeds advanced analytics. The key tasks within each stage are summarised in Table 2 below, together with the parties involved in each task.

Table 2: Key tasks and responsibilities

Stage	Tasks	Parties Involved*
Business requirements	Understand business objectives	ALL
	Agree requirements for application	ALL
Analytical solution	Understand data	BA
	Prepare data	BA
	Analyse data and develop models	BA
	Test and evaluate models	BA/Business
	Agree application development	ALL
Specify application	Understand operational requirements	BA
	Develop specification	BA
	Agree specification	ALL
Technical specification	Find the right NAG algorithms	NAG
	Agree on hardware/software requirements	NAG/Business
	Choose the programming environment (JAVA/ .NET/…)	NAG/Business
Develop application	Program and test	NAG
Monitor & evaluate	Plan initial deployment	Business/BA
	Evaluate results	Business/BA
	Plan roll-out deployment	Business/BA
Implement application	Train users	Business
	Make required system changes	Business
	Launch roll-out deployment	Business
	Monitor and evaluate results	ALL

* Key:

ALL = Business/BarryAnalytics/NAG

BA = BarryAnalytics

NAG = NAG Software consultants

Conclusions

Business users are demanding access to advanced analytics software, and seeking to embed this functionality into operational processes in order to achieve enhanced business performance.

The kinds of advanced algorithms that users require include forecasting models, predictive analytics and optimization tools. All of these are techniques from mathematics and statistics which have been extensively studied, and which are well understood by experts in the field.

Users can greatly benefit – in terms of development time, resource allocation and application reliability – by obtaining these algorithms from the NAG Library. This vast source of proven and tested routines provides all of the common analytical techniques - as well as those that are less common – for a variety of programming environments.

It is essential to take a joint approach when embedding advanced analytics into a business application – involving business knowledge, data mining expertise and software development skills.

Barry Leventhal is Director of BarryAnalytics, an independent data mining consultancy.

Stephen Langdell is a Senior Technical Consultant at the Numerical Algorithms Group Ltd, Oxford, UK.

References

1. http://smartdatacollective.com/douglautzenheiser/49914/gartner-says-predictive-analytics-are-hot-bi-topic
2. http://www.forrester.com/Srividya-Sridharan#/The+State+Of+Customer+Analytics+2012/quickscan/-/E-RES61433
3. http://smallbiztrends.com/2010/02/analytics-at-work-smarter-decisions-better-results.html
4. Aberdeen Group (March 2011) Embedding BI in Enterprise Applications: Magnifying the Analytical Impact, Research Brief.
5. http://dms.irb.hr/tutorial/dm_proces.php
6. http://www.nag.co.uk/market/articles/analytics_and_nag.pdf
7. http://www.nag.co.uk/numeric/numerical_libraries.asp