Deployment of Predictive Models

Sumit Kumar Bardhan

For more Information about Predictive Analytics Software, please visit our web site
http://www.predictiveanalytics.in or contact

Predictive Analytics Solutions Pvt. Ltd.
No.5, 1st Floor, 1st Cross,
P&T Layout, Horamavu,
Bangalore-560043

General notice: Other product names mentioned herein are used for identification purposes only
and may be trademarks of their respective companies.

SPSS is a registered trademark of IBM Corporation
Excel is a registered trademark of Microsoft Corporation

Deployment of Prediction Models by Sumit Kumar Bardhan

Deployment of Predictive Models
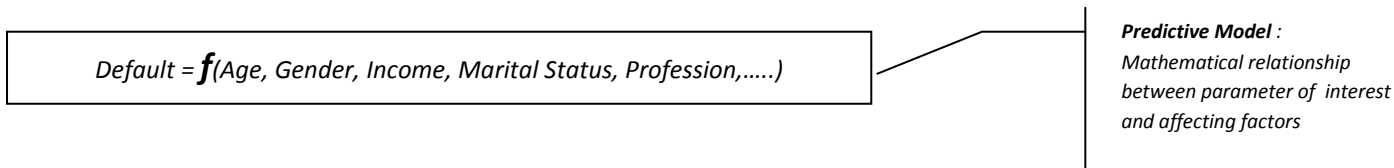
Predictive Models

In business and industry managers often need to predict the behavior of an important parameter that plays a vital role on the health of the business. These could include, among others, concerns such as,

- Identifying whether a particular loan applicant is likely to default
- Understanding the product purchase preference of individual customers
- Identifying the customers who are likely to leave the organisation's services
- Predicting the likely purchase spend for customers with given demographic profiles

Predictive models are built by observing the past behavior of the parameter of interest as well as the effect of those factors that affect its behavior, and then building a mathematical relationship between them. For example, in order to predict loan default, relevant historical on all the past defaulters would be studied in order to establish a mathematical relationship between defaulting behavior and other characteristics of the defaulters. This relationship is termed a Predictive Model.

Historical Information on Defaulters
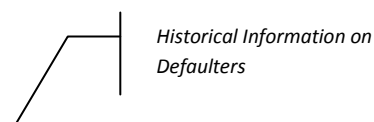
Parameter of Interest	Factors that affect behavior of the parameter					
Default	Age	Gender	Income	Marital Status	Profession	....
Y						
N						
N						
...						
...						
...						
...						

Default = f*(Age, Gender, Income, Marital Status, Profession,.....)*

__Predictive Model__ :
Mathematical relationship between parameter of interest and affecting factors

The name Predictive Model derives from the fact that given a mathematical relationship, and the values of the affecting parameters – often termed the **predictors** – it is possible to predict the value of the parameter of interest, which is itself called the dependent variable.

In simpler scenarios, a Predictive Model may be represented generically as,

$$y = f(x_1, x_2, x_3, x_4 .... x_n)$$

Where y is the dependent variable, i.e. the parameter whose outcome we are interested in predicting, and x_1, x_2, x_3, x_4.... x_n are the independent predictors, which affect the behavior of y.

Note that, in more complex cases the model may become more challenging, with the outcome of y becoming dependent not one, but on a system of equations.

Building Predictive Models

Predictive Models can be built using a variety of techniques, including

- **Multiple Regression**
- **Logistic Regression**
- **Neural Networks**
- **Decision Trees**
- **Cluster Analysis**

The choice of technique generally depends on the nature of the problem as well as the characteristics of the variable to be used in the analysis.

These techniques are available in Predictive Analytics software such as SPSS and R.

Scoring

Scoring is the method of substituting the value of the independent predictors in the mathematical relationship obtained by the modeling process, in order to obtain the 'prediction' for the parameter of interest. In the above example, newly observed values of the predictor variables *Age, Gender, Income, Marital Status and Profession* may be applied to the mathematical relationship, in order to obtain a predicted value for the parameter '*Defaulter*' .

Currently scoring may be done in about three ways

1. Manually input the values and calculate on pen and paper
2. Build the relationship as a spreadsheet formula, and the input the values
3. Run scoring in a batch mode in the software where the model was built in the first place, if this feature is supported

In the first two cases where there are significant challenges of computation and understanding of the theory behind the model, in the last case the scoring gets restricted to only those who have access to modeling software.

The Challenge of Usability

The model $y = f(x_1,\ x_2,\ x_3,\ x_4....\ x_n)$ is very powerful. For by predicting likely outcomes, it allows decision makers at all levels to decide their future course of action more objectively. However, given the skill levels required to actually use it, its use is not in the easy reach of those who can benefit the most, but yet do not have the necessary skills to understand complex mathematical

models. This is depicted in figure 1, which also shows that in comparison, the humble calculator, though not a very powerful tool, is nonetheless very easy to use.
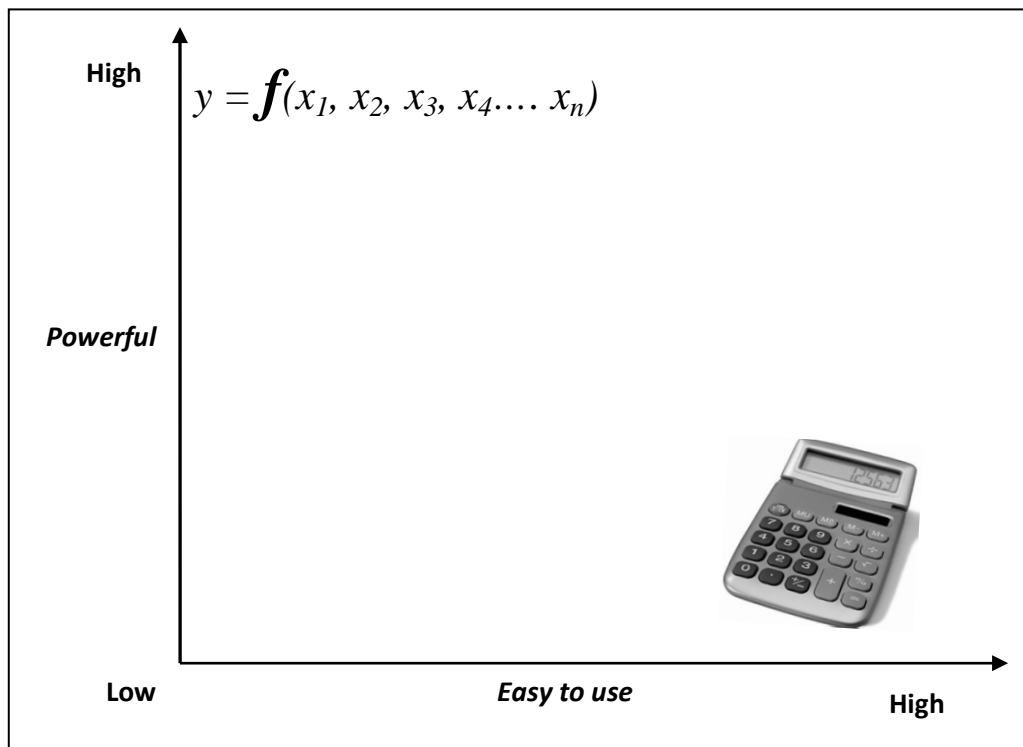

$$y = \boldsymbol{f}(x_1,\ x_2,\ x_3,\ x_4 \ldots x_n)$$

High

Powerful

Low   Easy to use   High

Figure 1. Power vs. ease of use

Deployment

In figure 1 above, can we shift the model to the right , so that while retaining its power , it becomes as easy to use as the calculator ? This is the challenge that deployment addresses.

In other words, deployment is the process of presenting a Predictive Model to the untrained end-user, so that while he or she can take full advantage of its power to predict outcomes, without an associated learning curve.

These are end users who need to predict, but

- Do not have any background or understanding of the mathematical principles of Predictive Models
- Do not have access to expensive analytical software
- Maybe geographically distributed
- May need frequent changes or updates to models as business parameters and priorities change

This can be achieved by translating the model via an application into an online form, somewhat like in figure 2, where the user inputs the values of the independent variables and the application returns the predicted outcome at the click of a button.

Figure 2. Online Prediction Form

Once we are able to derive this form from the Predictive Model, it in turn could be delivered to geographically dispersed users either via browsers (figure 3) or Smartphones (figure 4)

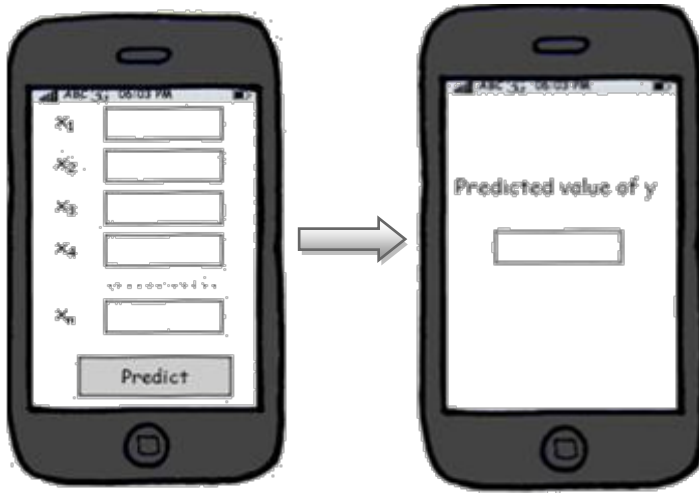**Figure 3. Online Prediction on a Browser**

Figure 4. Prediction on Smartphone

Once this is achieved we have an effective system for deployment that makes it possible for the untrained end-user to harness the power of Predictive Models as easily as a calculator, as shown in figure 5.
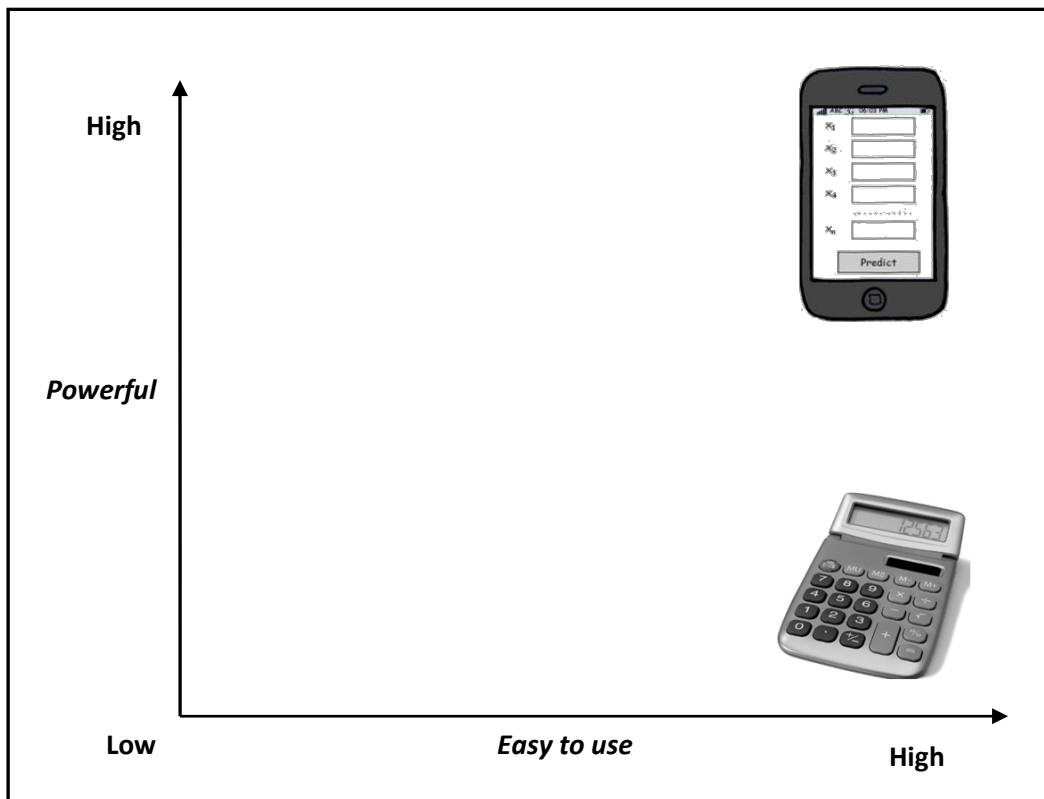

Figure 5. Power vs. Ease of Use, with Deployment

Deployment represents the last stage in the analytics process. With this in place, and the power of Predictive Modeling available to all users in the organization, Predictive Modeling can now be operationalised. It can be used in day to day operational decision making rather than being restricted in its use in an organizational niche for strategic purposes only.
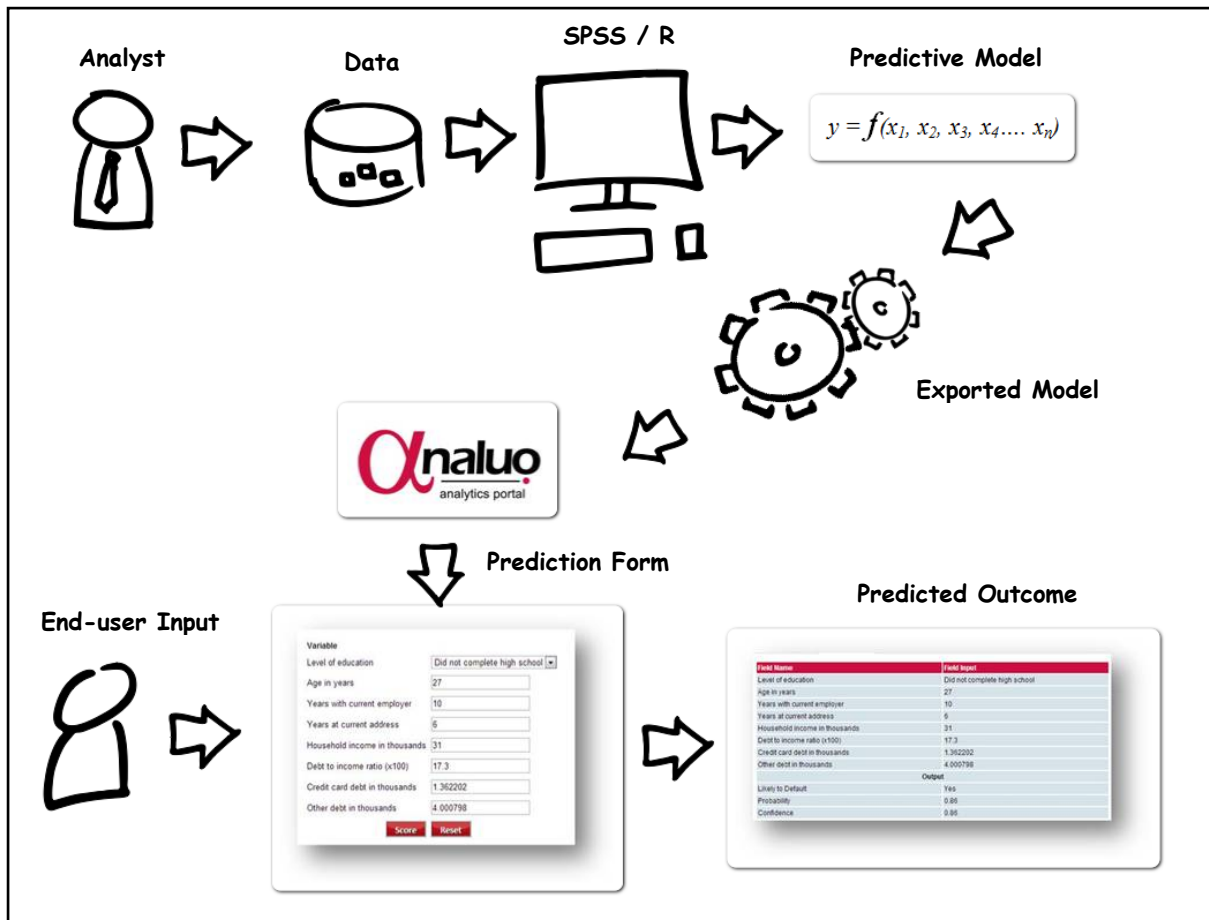
Deployment with Analuo

Figure 6 : Deployment with Analuo

Analuo is a deployment portal for analytics, intended to enable organizations in operationalising their Predictive Analytics initiative. With its patent pending technology for online prediction, addresses this challenge of deployment. As depicted in figure 6, an analyst uses data in SPSS or R to build a Predictive Model. This Model is now exported out as a PMML file. PMML, standing for Predictive Model Markup Language is an open standard defined by the Data Mining Group (*www.dmg.org*) for defining and sharing Predictive Models.

The Predictive Model, in PMML format, is uploaded to Analuo, which dynamically creates a interface, in the shape of a Prediction Form based on the model, which can is delivered either via a browser or on a Smartphone using Analuo's Android app. An end-user, even though not skilled in the techniques of Predictive Modeling, now has to simply enter the values of the independent variables or predictors in the form to obtain a prediction.