Machine Learning in Banking: How To Transform both Balance Sheet Management and Customer Services Provision

BY MOORAD CHOUDHRY

Advances in computer processing power have enabled established artificial intelligence techniques to become accessible to wider industry, which have now become a significant factor in the introduction of new and more sophisticated tools and platforms in today's global banks. In this article Professor Moorad Choudhry introduces machine learning and discusses how it's adoption can enable banks to optimise the "big data" that they already hold, allowing them to transform both their customer service provision and risk management practices.

In 2018 it is commonplace to suggest that "FinTech" and artificial intelligence are transforming the financial services industry and enabling firms to employ a powerful technology capability brought previously to the public's attention by Silicon Valley and the global tech companies. So called "challenger banks" and their digital mobile apps are disrupting the way finance works and the way customers conduct their banking. In truth, the possibilities offered by the intelligent use of available modelling techniques are immense, and the banking industry, be the high-street or "challenger" banks, have barely scratched the surface of what can be achieved by adapting their processes and customer service offerings through applying AI models.

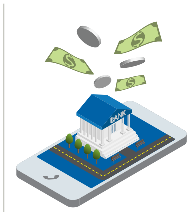
A particular branch of artificial intelligence known as machine learning, which is in fact an extension of existing disciplines including econometrics, multi-linear and non-linear optimisation, pattern recognition and computational statistics, has the potential to genuinely transform the way banks do business, in every aspect of their operation from customer front end to settlement back end. Some large as well as small banks are already employing machine learning in their risk management analytics processes, but the fact is that the technique can add value in any banking process, not just risk management and – equally significant – is a capability that is not restricted to only the large banks.

And therein lies the potential of machine learning for banks. Unlike certain prerequisites of the regulatory approval process for existing or new banks, such as information technology systems and human resources, not to mention the capital base itself, which are all extremely capital intensive and thus the preserve of large firms, implementing a machine learning modelling process in a bank's data analytics capability is not the sole preserve of established banks. In essence virtually any firm can set up and employ a model, assuming the modelling capability and fine tuning process of calibration is within the firm's intellectual capital base. But the power that the process brings can enable a bank to offer a truly transformational customer service experience and structure an optimised balance sheet. Adopting the technique allows any bank to be a genuine "challenger" and to represent the future of finance. (Of course, machine learning potentially also enables the established "high street" banks to further entrench their current market domination).

Machine learning models

One way to describe a machine learning model is as an extension of orthodox statistical techniques that seek to draw out relationships between different sets of data; for example, those between height and sporting ability or between interest rates and inflation. A well-fitted model will enable us to *infer*, or better still predict, an outcome based on the relationship between the independent variable (sporting ability) and the causal factor (height). But because the world and relationships within it are complex, (extending the analogy, there are many factors that drive sporting excellence, and as a causal factor height can be both a positive and negative driver), the more simple linear multi-factor econometric models are not generally of great value to someone looking to derive a deep understanding of a diverse data set.

Machine learning systems on the other hand, are designed around prediction, where identifying strong correlations between variables enables the user to move beyond the causal inference output of traditional econometrics. Hence they enable us to predict outcomes with much greater certainty, and to also understand relationships to a granular level. Combine this with the "Big Data" that a bank would already possess arising out of its ordinary business, and we have an analytical ability that enables the bank to understand its balance sheet, and the inter-relationships between individual customers, risk and regulation to a much greater degree than hitherto.



The basic concept

Machine learning models are many and varied. At one end are orthodox techniques from the well-established statistics and econometrics stable, such as multi-factor econometric models and logistic regression, while at the other end are more sophisticated models such as neural networks and random forests. Whatever their form, every machine learning model is comprised of the following elements: (i) the query or "problem" to be solved (in conventional econometric testing, the hypothesis); (ii) the data set; (iii) the model itself; and (iv) the optimisation algorithm. As part of the application of a model to a particular problem on a continual basis, there is also a fifth element which is the model validation and testing process.

In a banking context, the value of machine learning techniques lies in their ability to recognise general patterns and calculate approximate relationships in a data set, particularly for variables in the data in circumstances where no current analytical solution is available. Exhibit 1 shows sources of data that are well suited to machine learning analytics; note that the variables in such data sets would be considerable,

> The banking industry, be the high-street or "challenger" banks, have barely scratched the surface of what can be achieved by adapting their processes and customer service offerings through applying AI models.

In a banking context, **the value of machine learning techniques lies in their ability to recognise general patterns and calculate approximate relationships** in a data set, particularly for variables in the data in circumstances where no current analytical solution is available.

and standard linear econometric techniques would not be able to be as accurate a predictor of future behaviour of data compiled from such sources. Banks hold large volumes of data about all their customers, to an intimate and individual level of detail: pattern recognition of this data can improve greatly the efficiency of a bank's operations in many ways, which we explore later.

Exhibit 1 Examples of data sources that can be exploited using machine learning

Data Source	Examples
Online social media	Facebook, Twitter, Linked In
Communications	Phone, email, WhatsApp
Government	Central bank, Social security
Mobile app	Countless examples, with data trends in purchases, locations, habits, frequency, demographics, etc
"Internet of things"	Fitbit, smart meter, etc
Banking	Trends in: Payments, purchases, locations, value, age group preferences, entertaintment tastes, etc Financial products: Contact preference, frequency of renewal, retention, etc

Note: "Big Data" is defined here as large volume data with low information density

Machine learning models are either parametric or non-parametric. In the former type, the input data itself determines the values of a specified set of parameters; in the latter the input data influences the model's form, for example how many parameters are specified. It is rarely the case that only one model is the "best" available for any particular application, and different types of model may all perform satisfactorily in a given situation; however results and observation of their use should suggest which approach is the most suitable at that time and for that purpose.

Machine learning model form

In essence, a machine learning system seeks to address a problem that consists of two elements, the prediction and the causal inference component. A machine learning problem is formulated as an optimisation exercise, the solution to which determines the model parameters:

$$F(X,Y,\beta)\xrightarrow[\text{Optimisation Algorithm}]{}\beta$$

X and Y are the input and output variables respectively, and β represents the model's parameters, which are defined by the inputs and the form of the model. We optimise the objective function F(β) in order to estimate the model and describe the parametersβ. Data sets of the input data X and target variable Y are recorded in matrix form with dimensions of m×n and n×1 respectively, where m is the number of observations and n is the number of independent variables.

We can describe a learning system thus: assume that a known payoff function F is reliant on a policy variable X, an outcome Y and other control factors Z, which are independent of X and Y. Decisions with respect to X depend then on the total derivative:

$$\frac{dF(X,Y,Z)}{dX}=\left(\left.\frac{\partial F}{\partial X}\right|_{Y}+\frac{\partial F}{\partial Y}\frac{\partial Y}{\partial X}\right)\Big|_{Z}$$

Assuming we know Z, the unknowns in this expression are the prediction component Y and the causal inference component ∂Y/∂X. Assuming the policy variable does not impact the outcome (that is, ∂Y/∂X=0), this leaves only the prediction problem. Given sufficiently accurate knowledge about Y we can determine the outcome ("payoff") of policy action X. This is the space in which machine learning works. If we know Y, then we can determine the significance of the relationship between the outcome and the policy variable ∂Y/∂X≡β, where the coefficient β describes the strength of the relationship between the outcome and the policy variable. This is still within the econometrics discipline, but a machine learning model will also generate output for β. This is a critical quality

that means the system can in theory be employed by any bank, regardless of its history or the depth of its data set. (Of course, the values for β will exhibit bias so require care in their interpretation).

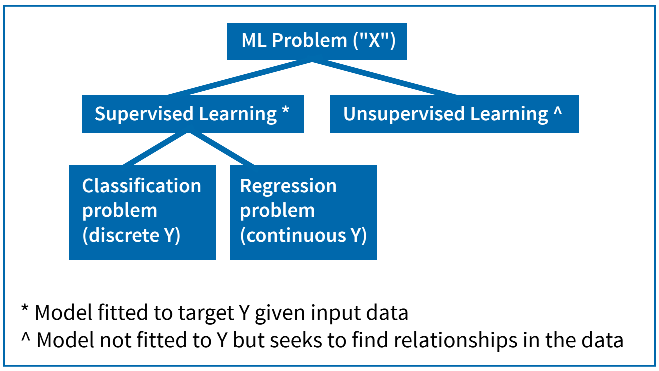
As we cannot assume that the two variables are known or are otherwise zero, machine learning techniques are best applied to "problems" requiring prediction and causal inference. As it happens, a large number of bank's daily processes and operations fit into just this category.

Defining the "problem"

Defining the problem is the key first stage of designing the machine learning system. Problems or *learning types* come in two forms, supervised learning and unsupervised learning. In supervised learning, the model is fitted to the input data given a target Y, while in unsupervised learning there is no target Y and the model seeks to identify patterns in the data. Supervised learning is closer to orthodox econometrics and would be used, for example, if forecasting GDP based on macroeconomic factor inputs. Unsupervised learning would be used, for example, if attempting classifications amongst a population based on their spending or entertainment habits. It is the unsupervised learning approach that has potential to bring considerable value to a bank's customer service provision. That said, both approaches may be used in the same context after patterns have been recognised (unsupervised learning) and the bank then wishes to predict customer behaviour and also forecast future business (supervised learning).

Exhibit 2 shows the categorisation of machine learning models. Supervised learning splits into classification and regression problems. In the former there is a discrete output variable (or variables) Y while the latter features a continuous output variable.

Exhibit 2 Machine learning problem types classification



As with any model, the efficacy of a machine learning algorithm is a function of its input data and the fit of the model. The model testing and validation process is an essential part of the implementation of a machine learning system. This will include the selection of data and continuous review of outputs to ensure validity of the model. (This process is analogous to the goodness-of-fit process in econometric analysis).

As with econometric models, **machine learning systems do not provide the "best" results on first use.** Regular adjustments to the shape (parameters) of the model and checking of the inputs and outputs are required.

As with econometric models, machine learning systems do not provide the "best" results on first use. Regular adjustments to the shape (parameters) of the model and checking of the inputs and outputs are required. Implementing a machine learning system requires data analytics training, model testing and validation. During this process the user calibrates the model so that it emerges with accurate generalisation characteristics. Evaluating the accuracy of the model (in supervised learning approaches) will involve comparing the model output with unseen target data Y.

In addition there is a trade-off between simple and more complex models. By definition a simple model (fewer parameters and fewer degrees of freedom) is quicker and cheaper to compute but may produce less accurate results. A complex model may provide better fit and more realistic predictive quality but require higher IT resource usage and hence be more expensive and time consuming. Determining the optimum model requires a trade-off in these areas.

The machine learning toolbox contains a wide range of techniques and methods. We list some of the more commonly used ones at Exhibit 3.

Exhibit 3 Machine learning models

Type	Techniques
Supervised Learning (Target Y given)	Decision trees Artificial neural networks Support vector machines k-nearest neighbours (non-parametric) Random forests (non-parametric) Naïve Bayes
Unsupervised Learning	k-means clustering Hierarchical clustering analysis
Recommender systems	Blend of different methodologies

Data set format

Traditionally machine learning models were used more for analysis of cross-sectional ("panel") data, but unsurprisingly when applied in the banking industry it is time-series data that is the most common data set that a problem is set around. This introduces additional potential issues in fitting the model and interpreting the results; but these are similar to what one must address in orthodox econometrics, such as missing relationships or trends in the data. Often when encountering possible errors in output, the solution is to break the time-series into shorter spans. Some machine learning approaches, such as neural networks, are more suited to time-series analysis than others, but the key again is to take one's time during the testing and validation process and also assess model efficacy by testing its predictive properties against actual historical observation.

Application and interpretation

As machine learning is in some forms an extension of traditional statistical techniques, in the finance industry it is unsurprising to see it has for some time been used for forecasting in markets and economics. A common use in central banks is as part of the inflation, unemployment and GDP forecasting process. But given machine learning is as much desired for its predictive qualities as its ability to quantify relationship parameters, there is a wider range of applications one can consider it for. For example, a seed-capital investor may wish to see which of a set of new entrepreneurial enterprises is most likely to "succeed", based on past and current patterns of success and failure in the same industry, geography and human capital sectors. A commercial bank may seek to determine what "optimum" mix of balance sheet assets and liabilities best meets the requirements of shareholder, regulator and customer, again based on past and current observations within the same peer group, jurisdiction and franchise.

These last two applications are examples of how "anomaly detection" can be of value in finance. This term is used to describe techniques that can identify within a dataset those observations that are outside of the established pattern or differ otherwise from the other observations in the data. A well-fitted model is thus able, for example, to assist the operational risk manager in a bank with identifying or better still predicting employee or customer fraud. We look at further practical applications in the next section.

Again, which machine learning algorithm type to select will be a function of the precise problem being considered, and more than one technique should be tested and compared before a final one is selected. When making the choice, it may not be immediately clear to the user how a model's parameter outputs should be interpreted, or how exactly the model generates its outputs from the input dataset. (Hence complex methods such as neural networks and decision trees are often referred to as "black boxes".) Interpretability may be an issue and is addressed during the validation phase; it is also minimised by implementing simpler rather than more complex models first, because small models are easier to interpret than large ones. As familiarity builds over time and the model is (where possible) layered with more complexity, the user can understand better a model's output and make more reliable assessments.

Machine learning, customer service and an optimised balance sheet

We have already introduced some common uses of machine learning techniques in banking and finance. A conventional application is in the field of risk management; for example (through anomaly detection), by predicting such things as loan defaults, employee or customer fraud. One can, for example, see how the former would assist greatly with IFRS9 compliance!

Certain machine learning models can still assist a "challenger" bank to target customers more accurately, because the model itself will calculate the parameters that enable predictions to be made.

However current applications (at least, those that are acknowledged publicly) are only the tip of the iceberg when considering the power of machine learning. And a key feature is that it is possible for any bank to apply the technique to both its front-end customer service processes as well as its back-end operational and risk processes. The tool is not the sole preserve of big banks.

Big data and customer service

A quote from Christopher Shallue, an engineer at Google, in *The Times* of 15th December 2017 encapsulates well the power of machine learning

> *"Machine learning really shines in situations where there is so much data that humans can't search it for themselves."*

This capability has led commentators to predict that many service functions that involve the review of extensive but standardised documentation, as carried out for example by solicitors and accountants, will soon be undertaken by a machine learning algorithm doing the same work in a fraction of the time. Banks hold considerable data on every one of their customers – their tastes and habits with respect to virtually every aspect of their lives. Applying pattern recognition and anomaly detection models to this vast dataset will enable banks to serve existing customers in specific tailored ways, as well as conduct marketing to prospective customers in a much more focused way than currently.

Consider these examples:

- Existing customers who fit a pattern that identifies a particular peer group, who have not purchased a financial product that their peers have, can be identified as predicted future buyers and marketed to specifically, using their preferred communications media (branch, smartphone, etc);
- Prospective customers who fit the pattern of specific peer group current customers can be the object of tailored marketing;
- Specific loan or deposit products that are popular with particular types of customer can be developed further (or introduced if the bank does not offer them);
- Specific customer franchises that are currently ruled out *en masse* (for example, non-bank financials, charities, offshore special purpose entities) can be considered individually based on patterns marking them out out individually as "safe", and hence worth doing business with.

These are instances of "unsupervised learning" problem pattern recognition, identifying patterns amongst customers with identical consumption habits, enabling banks to market to those who haven't bought their product yet. This is considerably more cost-effective than current approaches (such as postcode or area code targeting, or age group focus), which are often inaccurate or worse still, ignored completely.

Long data history further enables a bank to identify patterns and relationships and also trigger more accurate predictions; for example

Long data history further enables a bank to identify patterns and relationships and also trigger more accurate predictions.

While machine learning may not be the brave new world, it does offer a **new higher level in the bank strategy setting, marketing, risk management, and customer service process.**

some high street banks have over 30 years of data on their customers. But what if a bank does not possess 30 years data? Certain machine learning models can still assist a "challenger" bank to target customers more accurately, because the model itself will calculate the parameters that enable predictions to be made. Machine learning approaches add value irrespective of whether there is an excess of data or if there is very little of it. The key of course is in the modelling set up and calibration, but this is not an insurmountable issue. The implication for banks customer services provision is immense: specific product and customer tailoring, really knowing one's customer, whilst still retaining a human touch where it is desired by the customer, for the times and/or products when the customer wishes human interaction.

Big data and balance sheet management

Another aspect of bank's operations that would benefit from application of machine learning is balance sheet management. In the Basel III era a bank has three key stakeholders: shareholder, regulator and customer, and the asset-liability mix that best suits the requirements of each stakeholder differs in each case. Identifying the optimum balance sheet shape and structure demands an integrated approach to asset and liability origination, what the author terms "Strategic ALM", and seeing patterns and anomalies amongst the mix of product types and customer types would be invaluable at every stage of the business origination process, as well as the capital and liquidity planning process. For perhaps the first time in the history of banking, it is possible for a bank to truly structure its balance sheet in way it desires, and understand the multi-faceted nuances of the relationships between customers, products and market factors within it.

Conclusions

Machine learning is not necessarily the brave new world of finance, and it doesn't necessarily mean the end of one form of banking and its replacement by another. What it enables banks to become are masters of data analytics – their data. And genuinely understanding one's data in a bank is like fire for a stone age human, it enables one to conquer the environment. Of course, a multi-disciplinary approach to model implementation is needed. A machine learning model properly applied requires expertise in database management as well as algorithm building and statistical interpretation. It is a complex process but within the capability of any firm with the right vision and technical expertise.

Machine learning systems are not "new" as such, rather an extension of the tools used hitherto in econometrics. What is new is that large, multi-variate datasets can now be analysed to identify the deep relationships within them. Bank customers' patterns not just of their spending but in essence of their entire lifestyle can be assessed for an understanding of routine and outlier. Datasets originated from the balance sheet are precisely what machine learning systems are best capable of analysing, and a bank now has a means of modelling detailed and complex non-linear relationships amongst its customers, products and competitors. Hence, while machine learning may not be the brave new world, it does offer a new higher level in the bank strategy setting, marketing, risk management, and customer service process.

Professor Moorad Choudhry lectures on the MSc Finance programme at University of Kent Business School. He was latterly Treasurer, Corporate Banking Division at The Royal Bank of Scotland, and is author of *The Principles of Banking*.

Endnote

Note that "regression" has a different meaning in machine learning compared to its meaning in econometrics. In machine learning the word refers to a problem class, whereas in econometrics it refers to a type of model such as Ordinary Least Squares.