Natural Selection

Financial Services Forum

21-22 May 2018 • Hilton Sydney

Actuaries Institute

Machine learning in credit scoring

Ed Stokes

Contents

Credit Scoring

A BRIEF SYNOPSES

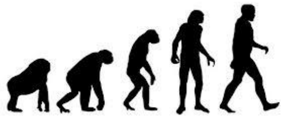
What are credit scores and why do they exist.

What is Credit Scoring?

- A discipline dedicated to guiding an organisation's credit decisions.
- Employs predictive modelling techniques to analyse historical data to quantify the likelihood of future events;
- An educated approximation of reality.

How did it all begin?

- Introduced by Fair Isaac and Company (FICO) in the US in 1956;
- In Australia externally developed scores are generally used as inputs to an organisation's internally developed credit score.



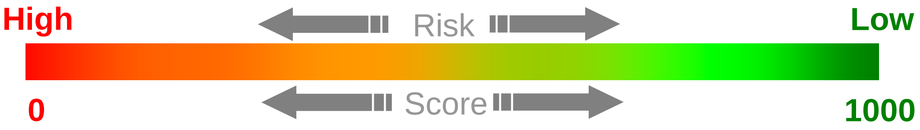
Who uses it and why?

- Banking and Insurance industry and more recently utility providers;
- Allows credit providers to centralise their decision making process and make more consistent and accurate decisions;
- Increases efficiency by automating processes.

How do credit scores work

Credit scores rank order risk

- Scores are a **numerical value** calculated either at the **point of application or monthly (i.e. behaviour scores)** which assesses the likelihood of the account defaulting some time in the future (typically 12 – 24 months);



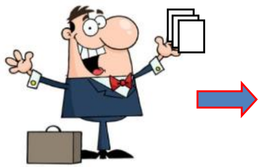
- A **low score** implies an account has an increased likelihood of **defaulting** on their commitment, conversely a **high score** implies it has an increased likelihood of **honouring** their commitment;
- Generally, any given score can be directly **mapped to a probability**, through a mathematical transformation;
- In general Credit Scores are often described as **a risk ranking tool**. That is, a lower credit score explicitly implies higher risk than a higher credit score;

Score Range	# Total	# Good	# Bad	Bad Rate
Min - <100	500	50	450	90%
100 < - 200	400	100	300	75%
800 < - 900	400	300	50	10%
900 < - Max	500	450	25	5%

Where do credit scores sit for decisioning

Scorecards are inputs into the credit decision framework

- Scorecards do not operate by themselves;
- Credit decision framework includes a combination of both **score** and **non-score** rules;
- Non-score rules are influenced by an array of strategies, such as other product performance;
- The final decision is based on the organisation's risk appetite to take on the risk.

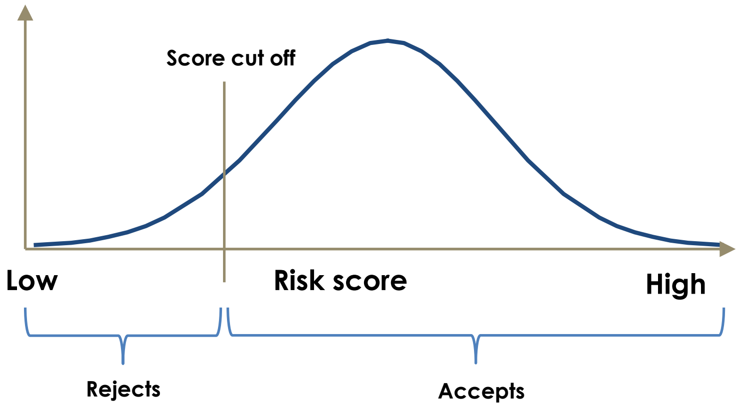


Credit decision rules

Credit strategies

- Age > 18?
- Insufficient surplus?
- Unemployed?
- Non resident?
- Irregular accounts?
- Adverse bureau?
- Impaired asset indicator?
- Bankrupt?
- Judgements?
- Poor score other product?
- P(D) rating?

Scoring



Final Decision

Modelling approaches

OLD WORLD VERSUS NEW WORLD

Rapid data evolution

Evolving environment

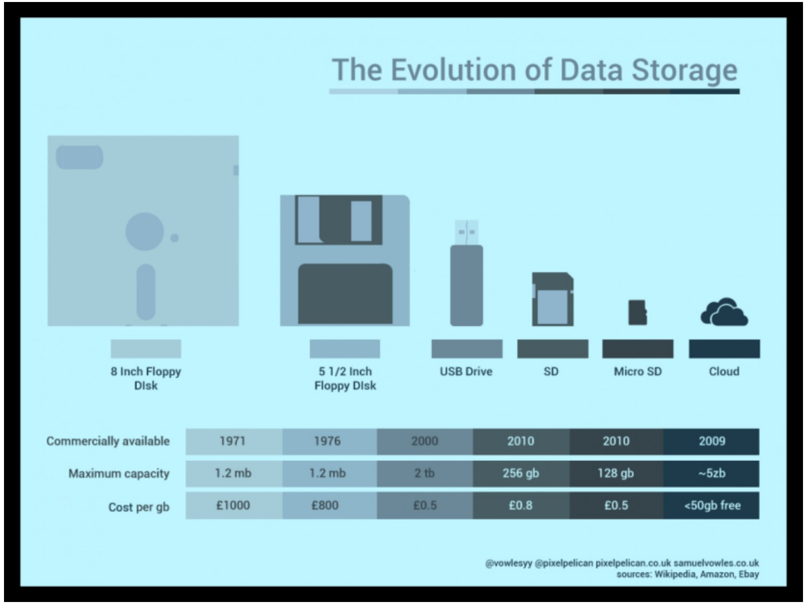
The enhancement of data storage and processing is enabling a **rapidly evolving** data landscape; allowing for new data feeds.

Increase discrimination can be achieved, if new sources of data are incorporated into credit decision models. Allowing for more informed credit decisions.

Accommodating new information is challenging, it requires continuous model rebuilds. If an **agile modelling approach** can be adopted, then a credit provider can take advantage of all sources of information as they become available.

Examples of new data sources:

- Comprehensive credit reporting "CCR";
- Transactional data;
- Open data.



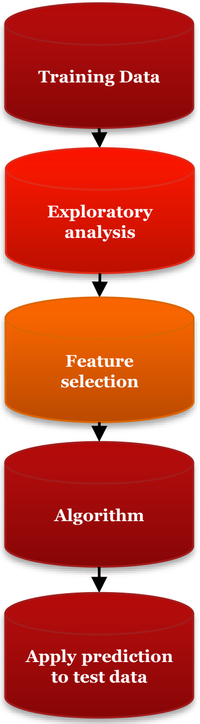
Current industry approach

Current industry approach – generalised linear models

- **Well-defined discipline:** Credit scoring has been a well-defined discipline for the best part of two decades, where generalised linear models 'GLM' have been the go to technique;
- **Data preparation is intensive:** the data preparation involved is a lengthy process, where data cleansing, feature engineering and stratification of the portfolio can take the best part of 6 months;
- **Long time to implementation:** the model build, implementation and sign off can take the best part of an additional 6 months. Taking approximately 12 months, in total, to build and implement models for a given portfolio.
- **Questionable relevance:** Due to the elapsed time, the data that the model is trained on may not be relevant to the current portfolio performance. Meaning, that sub-optimal decisions may be made.

N.B due to the lengthy end to end process, models will typically stay in production for 3 to 5 years, further exacerbating the relevance of the models parameters on new credit decisions.

GLM
Requires manual preparation of the data (e.g. discretising variables), this is resource intensive; stylistic, can be time consuming.
End model is rigid; i.e. set in stone, it is very hard to retrain the model quickly (i.e. will need to examine the impact on all of the variables in the model).

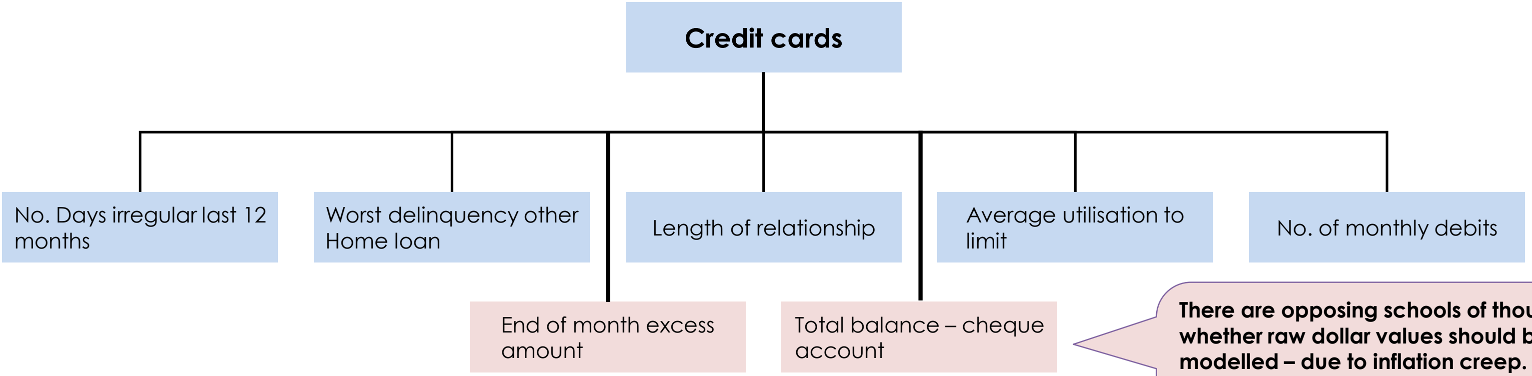


Current modelling method - GLM

Generalised linear models "GLM"

- A GLM typically has 6 - 12 predictive variables;
- N.B. a GLM is a parametric model – where a number of assumptions have to hold – this creates restrictions on how the model is built and often leads to a stylistic model;
- Once developed and implemented, these variables and their estimates are locked into a decision engine;

Example of variables in a credit card scorecard:



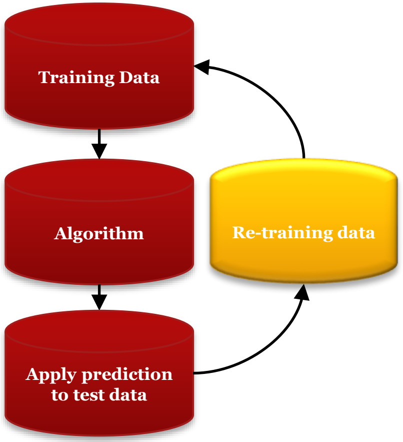
There are opposing schools of thought on whether raw dollar values should be modelled – due to inflation creep.

If an agile modelling environment were created – it would be a moot point.

Alternative industry approach

Alternative approach – gradient boosted trees

- **Less intensive data preparation:** A GBT requires significantly less feature engineering, due to its non-parametric nature, allowing a more seamless streamlined model development process;
- **Faster retraining:** Machine learning algorithms, like gradient boosted trees 'GBT', can expedite a model build and allow for quick retraining of the model;
- **Agile governance required:** A more recently trained model could only be implemented in a GBT world if an environment with agile governance is set up.



GBT
Automated feedback loop, whereby new results are fed back into the model algorithm to estimate a new prediction. There is minimal data preparation involved

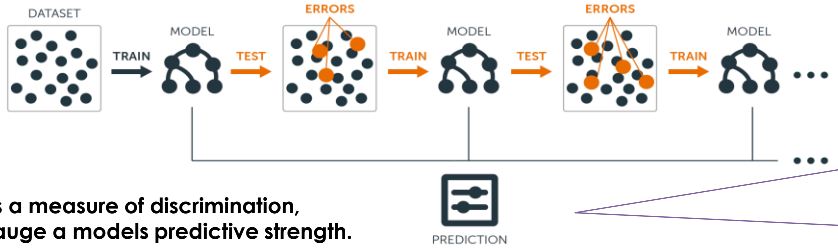
Alternative modelling method - GBT

Gradient boosting

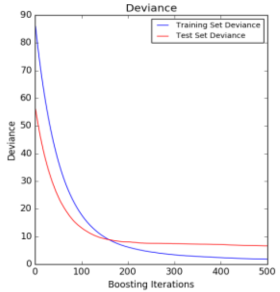
- A GBT is an additive regression model, which uses a sequence of regression trees to perform supervised learning;
- Boosted trees fit the training data over a series of sequential trees, where each iterative tree is regressed against the residuals of the previous tree;
- Each tree is constructed in a greedy manner, choosing the optimal split of the entry variable, based on the interaction of the previous variable to minimise variance.

Mitigate against overfitting

- GBT favours many shallow trees (e.g. depth of up to 3 - 6 nodes and typically "50 – 500" trees), this ensures robustness in the partitions;
- The maximum number of tree iterations is reached when the error deviance converges to a stable number (shown bottom right); this is achieved via measuring the loss of the training data against that of the cross validation data (essentially measuring the drop in gini*/gini equivalent between training and x-validation).



A GBT builds multiple "small" tree models off the error term of the previous tree. The node in each tree takes the optimal split of the entry variable, based on interactions with the previous variable.



***The gini coefficient is a measure of discrimination, commonly used to gauge a models predictive strength.**

Modelling approaches

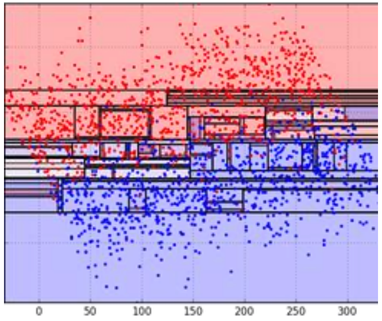
COMPARISON

GBT v GLM

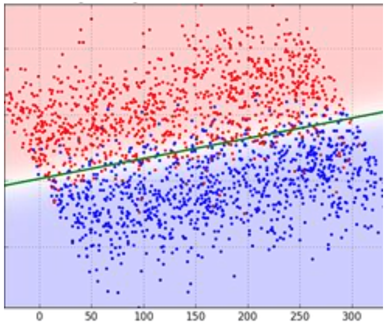
How does a GBT compare to a GLM?

- A GLM, simplistically put, **separates a region of space by fitting a line between two distinct categories**. This is advantageous when there is a simple target to fit the model too, and not many permutations/interactions within the data;
- A decision tree **separates different regions of space via creating strata** – i.e. isolates pockets by a myriad of interactive variables. This creates a natural segmentation process, whereby different weights are applied to different strata.
- N.B. a basic decision tree (e.g. CARTs) have a tendency to overfit the training data, as the model keeps drilling down to smaller strata to fit the model;
- Hence, Gradient Boosting mitigates against the tree's inherent desire to over fit the training data by building lots of shallow trees (i.e. 3-6 nodes), and offsetting each new tree against the previous tree's residuals.

Trees stratify



GLMs polarise/separates



GBT v GLM – the need for speed

Building a robust GLM is time consuming

- It takes multiple phases to develop a robust GLM, many of which are time intensive;
- ***N.B. not all bolded phases are required for a GBT build.***

Model steps	Description
Sample design	Remove observation that will create a bias
Variable reduction	Drop all variables that are single level or all null values
Automated bucketing of the variables	**Discretise the variables against the target (i.e. create univariate risk profile of the variable)**
Segmentation analysis	**Build preliminary models to see if discrimination is maximised**
Principal component analysis (variable clustering)	**Create family groups of like variables (i.e. similar risk angle); select variables from each family group with highest predictive strength**
Correlation analysis	**Remove variables with strong correlation and lower information values**
Manual classing of the variable	**Create manual buckets for final variables**
Build final model	Train model on the top 6 -12 most robust variables
Validate model on out of sample data (post development period)	Run model diagnostics to gauge model performance

Summary

Pros of a GBT relative to a GLM

- GBT can be **retrained on more recent data** – model is aligned to recent behaviours in the portfolio, this is particularly relevant with rapidly changing populations;
- Can rebuild models rapidly to **align to new strategies** and or to incorporate new information – e.g. CCR;
- Can trial new data sources easily – e.g. transactional data, to understand new risk relationships;
- **Significantly quicker to build** the model – e.g. less feature engineering (as the models objective function does this);
- Enhanced **discrimination** “potentially” – model’s ability to separate between future goods and future bads is greater.

Limitations of a GBT relative to a GLM

- **Requires an agile environment to implement** and monitor/validate;
- If models are updated frequently – will need to devise **new modelling diagnostic techniques** (e.g. typically models need 12 months of performance to test their discrimination);
- The model is opaque, regulator may deem to be **a black box**;
- Further to the above, **difficult to explain variable interactions** at a univariate level;
- The model is **prone to overfitting** if the correct controls are not in place.
- GBTs may exhibit a **faster decay rate** in discrimination – i.e. have a shorter shelf life.

Questions