HAPPY MEMBERS, SUSTAINABLE CREDIT UNIONS

THE OPPORTUNITIES IN CREDIT SCORING IN THE AGE OF BIG DATA AND PSD2

Fintech White Paper Series
December 2018

SOLUTION CENTRE

SOLUTION CENTRE

CONTENTS

Introduction

During the Renaissance in Italy, moneylenders would congregate at benches in city squares, their prevalence was such that the Italian for bench, '*banca*' lent its name to the English word bank. At this time, it was common for moneylenders to go out of business. In this event, they were required to ceremonially break their bench to signify they were no longer trading, the Italian for this, '*banca rotta*', literally 'ruptured bench' giving rise to the English – bankrupt.

Credit has long been a key driver in any economy and access to it is a key leveller in society, making home ownership or the purchase of everyday or exceptional needs a reality for people who could not otherwise afford them. However, moneylenders need to limit their risk, accordingly, the access to credit is not a right and lenders must devise strategies for choosing the applicants with the best ability to repay so that they can avoid a *banca rotta*.

At its core, credit scoring is the science of determining who can and cannot have access to credit. This is achieved by applying rules and models in determining credit worthiness. However, credit worthiness is relative, not an absolute; for example, a person may be deemed good for a secured loan but not for an unsecured because the risk is lower. These broader considerations of credit risk management form part of credit scoring and will be explored in this White Paper together with how the scoring technologies have evolved over time.

Speaking to the American Bankers Association on the 7th October, 2002, then Chairman of the U.S. Federal Reserve, Alan Greenspan said[1];

"*Credit-scoring technologies have served as the foundation for the development of our markets for consumer and mortgage credit, allowing lenders to build highly diversified loan portfolios that substantially mitigate credit risk. Their use also has expanded well beyond their original purpose of assessing credit risk*."

Continuing, Greenspan made an observation that is all the more profound when considered in the context that within six years of giving this speech, a subprime mortgage crisis would lead to the collapse of Lehman Brothers, precipitating the great recession;

"*The use of credit-scoring models has taught bankers - sometimes through costly experience - the value of continually updating the database on which the model operates. Indeed, one can speculate that some of the problems this year in subprime credit card losses may well represent an insufficiently long data series to score successfully such credits during a recession. The experience with credit-scoring models underlines the necessity of basing more-sophisticated quantitative approaches, approaches that seem to have served the banking system so well when applied initially, on a longer and larger database of loss experience.*"

In this speech, Greenspan touched on a number of topics which will be covered in this White Paper, including; the importance of quantitate data in credit scoring, the evolution of credit scoring technologies to serve other purposes and lessons learned from past mistakes in credit scoring. The advent of 'Open Banking' under PSD2 means that credit unions may have greater access to quantitate data than ever before and that will also be explored.

[1] https://www.federalreserve.gov/boarddocs/speeches/2002/20021007/default.htm

Part A – Background

The basics of credit scoring

Credit scoring is the use of statistical models to transform relevant data into numerical measures that guide credit decisions

As can be seen from this definition, credit scoring always relies on the input of relevant data. Quite simply, the more data the better in terms of model accuracy. Data can come from a number of sources, including;

- Payslips or social welfare receipts
- Bank statements
- Loan or credit card statements
- Internal data on past loan history
- Credit reference agencies
- Court records
- Transport records (e.g. is car being bought under finance, an import, NCT, etc)
- Social media profiles[2]

In credit scoring terminology, where a lot of data is available the file is "thick". By contrast a "thin file" is one where there is little or no data on which to base a decision.

Once data is collated the credit scoring may begin. Chapter 3 of the third FinTech White Paper was an introduction to the topic of credit scoring. As was noted there, credit scoring varies in complexity but typically involves 3 component parts;

BUSINESS RULES	SCORECARDS	DECISION ENGINES
Inbuilt rules to ensure that lending occurs in conformity with risk management policies. Compliant deals can be processed automatically and exceptions reported in real-time.	Multiple data points, weighted based on their importance, which when totalled provide the credit score. Each product will have its own scorecard, perhaps two based on the availability of data (so called thin and thick file)	Software which collects data, analyses it and creates a score in accordance with the scorecard. Pre-set levels determine if the request is automatically approved, declined or sent for review.

We'll explore each of these areas in greater detail in this White Paper, however, credit risk management considerations on loss severity like Expected Loss, Loss Given Default and Exposure at Default are out of scope.

[2] Behavioural credit scoring analyses data including social media activity to run psychometric tests to predict a person's credit worthiness. This remains a controversial approach nut Lenddo is an interesting case study.

The origins of credit scoring

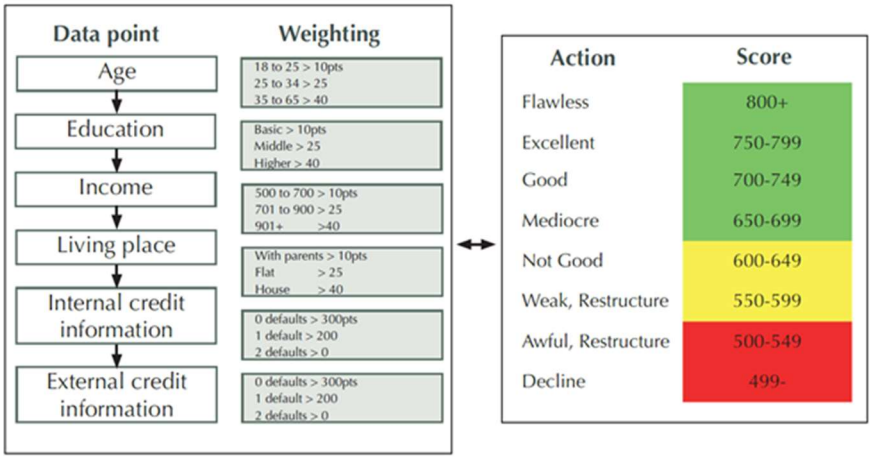
Traditionally, loan underwriting was based on the 5 Cs;

Character, Capacity, Capital, Collateral, Conditions

These were mostly subjective tests, an experienced underwriter applied these assessments, often having local knowledge of the applicant to assist in making a determination.

As banking evolved, decision making became more centralised. Customers were also more mobile as society changed and people moved for work or education. The traditional relationship-based assessments became less reliable so many banks shifted their focus to more objective tests, so that where they did not have a pre-existing relationship with an applicant, they could apply a rational set of rules in order to assess credit risk and make a lending decision[3].

By the early 1940s, some bankers were beginning to experiment with models to predict the credit risk of an applicant based on data regarding age, gender, occupation and assets. These early models were based on research published by English statistician Sir Ronald Aylmer Fisher on a technique he called 'linear discriminant analysis'. Peculiarly, this technique was initially employed to identify species of iris flowers but is today recognised as the genesis of modern credit scoring[4].



The rise of computerisation throughout the 1960s and 1970s allowed for the automation of the computation of large data sets in credit scoring. In the early days however, the power of these computers was extremely limited. For example, the IBM 7090 mainframe computer of the early 1960s could handle 25 variables for up to 600 applicants at any one time[5].

Over the coming decades there were massive improvements in computing technology. Anderson observes, this made it increasingly possible for lenders to invest in data warehousing, data mining and inhouse scorecard developments[6].

---

[3] Many credit unions continue to enjoy long standing relationships with their Members and this provides a distinct competitive advantage

[4] Johnson R.W. (2004) Legal, Social and Economic Issues in Implementing Scoring in the United States

[5] Myers J.H. and Forgy E.W. (1963) The Development of Numerical Credit Evaluation Systems

[6] Anderson R. (2007) Theory and Practice for Retail Credit Risk Management and Decision Automation

Part B – Credit Scoring in Practice

Business Rules

These will typically be a common feature in credit unions, ensuring standard rules are applied in the review of each application. Generally, these rules are designed to verify that an application falls within the credit policy and/or credit risk appetite statement, but also allow for the structured application of the 5 Cs of underwriting. Rules may include;

- Has past loan history with credit union
- Has saving history with credit union
- Has savings balance to attach against loan
- Is in employment
- Has been with current employer for greater than 6 months
- All other debt calculated in Debt Service Ratio
- Net Disposable Income remaining after new loan repayment exceeds minimum

Often, business rules will be used to expedite certain applications because they can easily be classified as low risk. The importance of lean processes in credit review is increasingly important as credit unions compete directly with banks on speed of turnaround. This is explored in greater detail in the Digital Transformation Programme and the Hive platform makes use of certain business rules based broadly on the 5 Cs to allow for the fast tracking of certain low risk files.

The Affordability Calculator developed under digital marketing uses business rules for key financial metrics supplied by credit unions to calculate an applicant's affordable monthly repayment and reverse calculate their max lend facility. By deploying this calculator on websites and landing pages under digital marketing we can tell an applicant how much they could borrow up to. This has an immediate impact as average application value increase by circa. 20% once the calculator is launched and term applied for increases by almost 12 months. By increasing the average value and term on applications it has greatly reduced the volume requirement in order to hit loan growth targets.

The SAM platform developed as part of the mortgage framework is also an example of the structured application of business rules in mortgage credit assessment. The main features here are the application of key financial ratios determined by a credit union as their risk appetite together with extensive checklists to ensure non-financial business rules are satisfied by an applicant. The system automatically calculates the key ratios fall within credit policy and reports on the checklist items completed by the underwriter.

Depending on the exposure to the credit union in a default event, the exceptions reporting on any file not satisfying all business rules may be onerous. For example, the exceptions report included as part of the cover note on any mortgage file not satisfying all business rules should be very detailed and set out the underwriter's rationale for recommending the file for approval despite all rules not being satisfied. Generally, all exceptions should be referred to the Credit Committee for review.

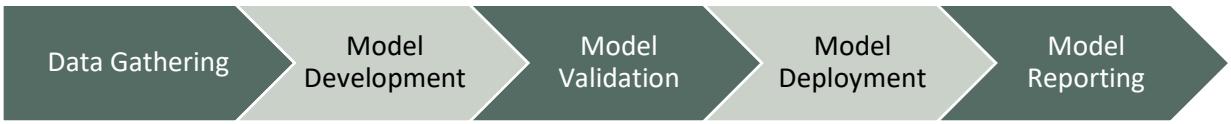
Credit Scoring

"… too many financial institutions simply outsourced their risk management. Rather than undertake their own analysis, they relied on rating agencies to do the essential work of risk analysis for them."

Lloyd Blankfein, CEO Goldman Sachs
(Financial Times, 8th February, 2009)

The 2008 financial crisis served as a reminder that it is incumbent upon every lender to take responsibility for credit risk management. Credit Scoring offers many benefits in assessing the risks of loan applications and the inherent risk in the portfolio under management, but it is important that senior managers and directors are aware of the basic principles underpinning Credit Scoring modelling, that these are kept under review and adjusted where needs be so that they remain consistent with the risk appetite of the credit union.

By far the most common scoring model still employed by lenders are scorecards. There are most basically 5 stages involved in developing and managing credit scorecards[7];



Data Gathering

As was noted previously, the single most important factor in the performance of credit scoring is the quality of the input data. Historically, the data that was gathered was largely qualitative in nature and took place in a face-to-face meeting between an applicant and underwriter. Today this is still recognised as one of the most reliable methods for credit decision making as an experienced underwriter can ascertain that an applicant satisfies the 5 Cs and is a reasonable prospect to repay a loan. As lending became increasingly computerised, local knowledge became less reliable and regulatory requirements like Basel II were introduced, many financial institutions became increasingly reliant on analysing quantitative data. This becomes increasingly important as financial institutions embrace digital transformation and explore methods to achieve automated straight through processing of applications.

For the most part, credit unions have had many challenges in obtaining quantitate data and have been largely restricted to that contained in their own banking system on loan and share activity and more recently the ICB for repayment histories with other financial institutions. Bank statement collection has heretofore been a very manual exercise, although contained within them are perhaps the richest set of quantitative data available for underwriting. Open Banking under PSD2 is intended to digitise the sharing of banking records between banks and other lenders with the intention of increasing competition for the benefit of consumers. There are many opportunities in this and this will be explored later in this Paper. Hive has also been developed to allow for the structured sharing of 'Meta Data' at a group level so that data on credit risk can be analysed to take advantage of the law of large numbers for risk probabilities.

[7] Siddiqi N. (2017) Intelligent Credit Scoring (2nd ed.)

Scoring Development

Many credit unions will choose to purchase credit scoring models so this section merely seeks to provide a basic overview of credit scoring to assist in reviewing options.

Credit scoring is the use of predictive models (algorithms) to rank cases by their probability of being 'good' or 'bad' at a future date based on a lender's past experiences. The most common algorithms for scoring are logistic regression formulae used by scorecards, for example;

$$Logit\ (p_i) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$$

Where:
p = outcome, usually a probability of a binary outcome for example 'good' or 'bad'
β = intercept of the regression line
x = an input variable

In practice, what is being collected is a series of scores based on an applicant's circumstances under a list of variables. These scores are based on a lender's past experience. For example;

Variable		Score		Score		Score
Age	18-30	1	31-40	2	41-50	3
Years with current employer	<2	1	3-8	2	8-15	3
Past arrears	0	3	1	-1	2	-3

This is just a brief example, in practice the scorecard will rank an applicant under a long list of variables which may include;

- Age
- Home – owns, rents, lives with parents
- Years at current address
- Employment status
- Public/private sector/self employed
- Years with current employer
- Work telephone number supplied – yes/no
- Past arrears
- Education
- Number of credit cards
- Has overdraft – yes/no
- Source of loan enquiry – in branch, online, by phone

Once a score for the applicant, given their personal circumstances is obtained, the algorithm will calculate a probability of default. Generally, this will fall within a range where the application is accepted, rejected or referred to a more senior underwriter or the credit committee.

Part C – Future Developments

Big Data

There are 2.5 quintillion bytes of data created each day and the growth rate is such that 90% of the worlds data has been created in the past two years alone[8]. To put that number in some perspective, that amount would fill 10 million blue ray discs which if stacked on top of one another would be the height of 4 Eiffel Towers (or 10.7 Dublin Spires) stacked atop of one another.

The use of data by financial institutions to ascertain credit risk is not new, credit bureaus have operated and collected data on customers for well over a hundred years. Where big data is impacting the evolution of credit scoring is that more structured data is available from more sources and computer technology has advanced to such a degree that this data can be analysed with more sophisticated algorithms.

Big Data is characterised by the 4 Vs;

1. Volume – the sheer amount of data generated by individuals each day
2. Velocity – the speed at which data can be collected, stored and analysed
3. Variety – the availability from multiple sources, no longer reliant on credit bureaus
4. Value – the ability to convert data into positive business outcomes

One case study that may point to a growing trend by financial institutions to harness big data to innovate in credit scoring is Lenddo, a Singapore based credit scoring company that analyses data from non-traditional sources like an applicant's social media and smartphone usage.

Lenddo launched first in the Philippines in 2011 as their founders recognised that one billion people in the developing world were moving into the middle class and would need access to credit, however, their 'thin file' or lack of credit histories would prejudice them under the traditional scoring models employed by many financial institutions. Primarily, their software is still used by lenders in the developing world to score applicants from the emerging middle class for small loans, although, they have recently launched into developed economies also.

One method they use is for an applicant to invite friends and family to join what Lenddo call their "trusted network" on social media. The software then uses an algorithm to analyse and evaluate the applicants trusted network. This is combined with any data available from traditional sources together with a psychometric analysis from the application form to develop a credit score[9].

It should be noted that this method of credit scoring remains controversial and that there is limited independent research to verify its effectiveness. Some commentators have also raised concerns that the method prejudices socially isolated people.

---

[8] https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#421aa05a60ba

[9] https://www.lenddo.com/pdfs/Lenddo_FS_CreditScoring_201705.pdf

PSD2 Open Banking

One of the more promising developments from a credit union perspective and the ability to access data for credit scoring will be open banking under PSD2. While a thorough examination of the Regulation is outside of the scope of this Paper, for present purposes it is sufficient to summarise that in order to increase competition for the benefit of consumers, banks offering current accounts will be required to provide APIs to other lenders so that customers can instruct the sharing of their financial data and avail of the best offer available to them. There are many other facets to PSD2 but our focus here is the opening of banking platforms.

Open Banking is a UK initiative[10] that is well advanced in requiring banks to provide APIs so that customers can share their data with other lenders to compare credit options. Ireland is slowly catching up however, Ulster Bank have published an API and more recently AIB have also made their API available[11].

In practice, for a customer filling in an online loan application, they will be asked if they would like to contact their bank to instruct them to supply their financial data for the purposes of underwriting. Many API intermediary companies have been established and will white label this into the online loan application form;

From here, a customer can use their online banking login details to authorise and instruct the sharing of their financial data from their current account provider to another lender.

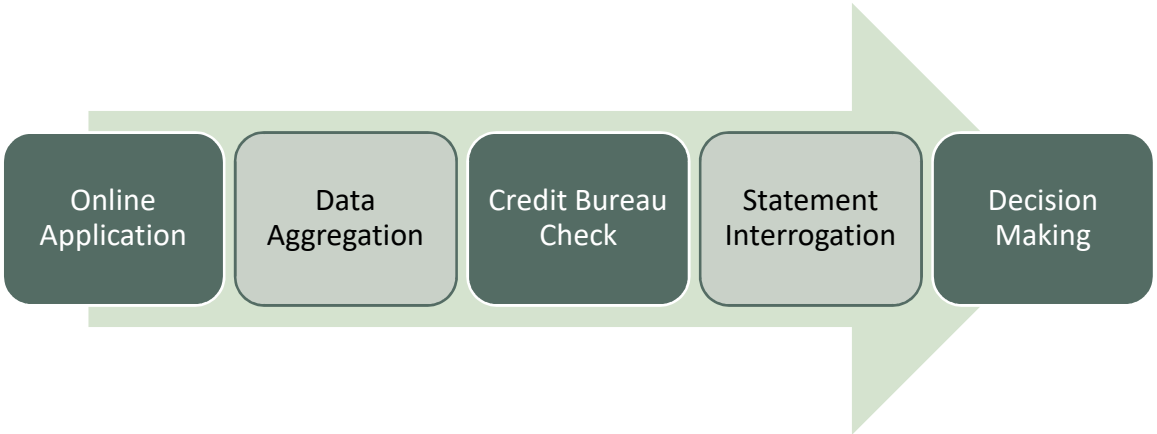
Once this data is shared in digital form, companies such as Nordigen, who presented at the Solution Centre FinTech Conference in 2017 can be used to analyse the transaction data contained within the bank statements. Their algorithms will analyse an applicant's spending patterns and flag any suspicious or undesirable spending (e.g. gambling) to perform credit scoring. These algorithms, analysing large quantities of data are in their infancy but are already outperforming traditional credit bureaus.

[10] https://www.openbanking.org.uk/customers/what-is-open-banking/
[11] https://www.siliconrepublic.com/enterprise/aib-open-apis-psd2

Conclusion

Credit scoring is a vital component in the credit risk management cycle for many financial institutions, however, the financial crisis served a stark reminder for many lenders that it is never wise to outsource responsibility for risk management. Where external vendors are used it is important that senior management take the time to understand the basic principles underpinning the credit scoring methodologies to be in a position to challenge any anomalies. Continuous validation of scoring models is also required to ensure they continue to perform as expected as financial cycles shift.

As Big Data is increasingly available, credit scoring methodologies will evolve in their sophistication to harness ever greater insight to an applicant's probability of default. For credit unions to avail of the opportunities presented in accessing and employing Big Data to mitigate credit risk, speed up decision making time and facilitate greater automation in the lending process will require an alignment of multiple systems from multiple vendors. Broadly, the process will follow the following flow;



Hive has been developed so that it may act as a base platform for the integration of multiple solutions to achieve straight through processing. It has also been built on a multi-tenancy environment and this serves two purposes – integration costs are shared collectively and meta data can be shared to improve feedback to the chosen solutions.

In any event, the ability to lower costs of delinquency and in turn achieve lower cost of credit for Members while expanding the ability to approve applicants who have a thin file should not be ignored. The technical challenges in implementing credit scoring, particularly as that sector undergoes significant disruption over the coming years should not be underestimated. Once launched, credit scoring will also require managers and directors to learn the mechanics of how it operates, have the confidence to challenge the assumptions used in modelling and constantly validate the models to ensure they remain accurate and reflect the credit unions risk appetite.

APPENDIX I – Other Titles in This Series

HAPPY MEMBERS, SUSTAINABLE CREDIT UNIONS

The *Happy Members, Sustainable Credit Unions* series of White Papers have been published by the Solution Centre to explore topics around Financial Technology ('FinTech') and Digital Transformation.

FinTech projects need to be viewed in their totality rather than in isolation. Failure by most financial service businesses in introducing FinTech solutions or Digital Transformation are due to the adoption of a scattergun approach resulting in processes and technologies that do not work together. This series of White Papers affords us an opportunity to explain how individual projects in a much broader programme of work fit together so that we can avoid those pitfalls.

Intended to be more useful than academic or research papers, there is a specific focus throughout the series on practical considerations and implementation.

The Opportunities in Fintech

Volume 1 (January 2017)

Volume 1 is an introductory text to give readers a foundational understanding of some basic concepts and terminology relating to FinTech. It examines the opportunities and challenges facing credit unions with the rise of financial technology together with target operating models. Volume 1 also includes a glossary of common terms, phrases and abbreviations relating to FinTech.

Volume 2 (April 2017)

Volume 2 focuses on practical examples of how FinTech drives deeper insights, improves member engagement and experience, back office efficiency and governance. It examines best practice in the FinTech and Credit Union segments, the degree to which we can trust machine decision making and how to avoid common mistakes.

Volume 3 (September 2017)

Volume 3 looks at the practical implementation of FinTech solutions and was the foundation in advance of the Business Case and Implementation Plan for Digital Transformation (June 2018). It covers data validation to drive business intelligence and in preparation for artificial intelligence solutions, appropriate KPIs for financial institutions in the digital age and scopes straight through processing of loan applications.

The Opportunities in Online Chat (April 2018)

This White Paper explores why Online Chat is expected to become an important engagement channel for most businesses with a specific focus on the successful implementation of Online Chat solutions.

The Opportunities in Point of Sale Credit and Offsite Engagement (Dec 2018)

This White Paper explores how technology can be leveraged to support lending in a retail environment at a point of sale, but also to proactively engage with members 'in the field' or offsite.