REPRINT

THE NEW AGE OF CREDIT SCORING AND DECISIONING

REPRINTED FROM:

RISK & COMPLIANCE MAGAZINE

JAN-MAR 2019 ISSUE

www.riskandcompliancemagazine.com

Visit the website to request
a free copy of the full e-magazine

Published by Financier Worldwide Ltd
riskandcompliance@financierworldwide.com

www.riskandcompliancemagazine.com

MINI-ROUNDTABLE

THE NEW AGE OF CREDIT SCORING AND DECISIONING

PANEL EXPERTS

Naeem Siddiqi
Director, Credit Scoring and Decisioning
SAS Institute Inc.
T: +1 (905) 471 2689
E: naeem.siddiqi@sas.com

Naeem Siddiqi has advised and trained bankers in over 20 countries on the art and science of credit scoring. He has worked in retail credit risk management since 1992, both as a consultant and as a risk manager at financial institutions. At SAS, he played a key role in the development of several products relating to credit scoring. He is currently responsible for managing SAS's end-to-end credit scoring and decisioning solution.

Nikolay Filipenkov
Principal, Credit Scoring and Decisioning, EMEA
SAS Institute Inc.
T: +7 903 728 9245
E: nikolay.filipenkov@sas.com

Nikolay Filipenkov is part of SAS risk business consulting and leads the credit risk modelling and decisioning initiative in EMEA. With over 15 years of experience in machine learning and 13 years in risk management, he has been a practitioner as both a banker and consultant. Before taking the current role, he was head of the risk management practice in SAS Russia & CIS for five years. In this role, he has designed and managed projects for banks, insurance firms and corporate clients in credit risk, decisioning, Basel implementation, fraud prevention and operational risk.

Abdullo Akhadov
Regional Head of Credit Risk Modeling, Machine Learning & Decisioning
SAS Institute Inc.
T: +65 6398 8963
E: abdullo.akhadov@sas.com

Abdullo Akhadov is regional head of credit risk modeling, machine learning (ML) & decisioning covering the APAC region, where he advises financial institutions on topics related to corporate and retail credit process, credit risk modelling, FinTech, risk and technology strategy, risk governance, IRB and ICAAP implementation, data quality, credit portfolio management and the application of advanced analytics in credit risk management. With over 13 years of experience in credit risk management, he has held various senior management positions covering business and technology aspects of credit risk management across several multinational banking groups.

R&C: In what ways has credit scoring changed in recent years? What factors now drive the decision-management processes of lending organisations?

Siddiqi: Over the last 15 to 20 years, credit scoring has become more analytical and mathematical. There used to be many business analysts building models, but now it has become more specialised, with decision scientists taking on this responsibility. This has partly been driven by a high demand for credit scoring and modellers globally. Of course, there has also been the Basel Accords (BASEL II) and the IFRS 9 International Financial Reporting Standard, as well as various other regulations that prompted many banks, even smaller ones, to develop their own models internally.

Filipenkov: Credit scoring has dramatically changed in recent years. The availability of new data sources significantly increased the capabilities of the machine learning (ML) techniques that financial institutions can use in the lending process. The decision process is on the one hand becoming much more complex, involving more models, and on the other has become much faster due to market demand and the increasing accuracy of automated decisions. In addition, decision management and model risk management has become more complex, requiring an agile approach. New risk or marketing models are being introduced and modified every day. There may be thousands of different decision paths depending on customer segmentation to deliver customer-focused offers. This means that the model and decision management infrastructure should support this complex environment, and allow rapid changes, testing and deployment of new models and decision logic.

R&C: How has the current technology changed the way organisations provide credit decisions? Have consumer expectations changed?

Filipenkov: The swift pace of technology has made it easier for organisations to get new data, process this data using ML techniques and provide smarter and more focused credit decisions. Many organisations are now using data from mobile applications and Facebook posts, as well as telecom data which allow them to provide credit decisions faster and more accurately. Consumer expectations have changed also. Customers expect a financial institution to answer them immediately once they submit an application. They also require personalised offers which consider their previous interaction with the organisation and which address their current needs.

Akhadov: The world is changing. In the last decade there has been explosive growth in technology adoption. This has been fuelled by technological breakthroughs and the development of technology infrastructure. Global development and decreasing margins pushes financial services organisations to look for new customers in previously under or unserved customer segments. We have also seen significant generational differences in values and preferences. Over the past 10 years we have observed dramatic changes in consumer needs, behaviour and expectations. Consumers are now seeking easy access, choice, better control and speed. They also expect that we know everything we need to know about them and can provide relevant and 'on-time' offers and decisions. Moreover, in this digital age, we have been conditioned to expect an immediate and relevant response, such as immediate loan approval and interest rates adjusted to personal risk levels. While traditional financial institutions are mostly reactive and conservative in their offerings and processes, new entrants, such as FinTechs, are closing existing gaps. FinTechs are the ones who mostly drive changes and disruption in the area of credit scoring at the moment. Their focus on customer experience, use of the latest technologies, such as advanced ML algorithms, combined with cutting edge alternative and trended data for credit scoring, has helped them become leaders in the personal loan industry. Both groups – traditional organisations such as credit bureaus, banks and small money lenders, and new entrants like e-commerce and P2P and FinTech lenders – are utilising non-traditional data and advanced ML algorithms to include those who lack traditional data and those which, due to the limitations of traditional models, were previously considered bad customers.

Siddiqi: Consumer expectations have changed substantially. Speed is the big factor now. In the past, because application paperwork was processed manually, customers might wait two or three weeks before receiving a decision on a mortgage or a car loan. Decisions are now expected to be instantaneous and that has given rise to FinTechs challenging banks for their lending book. The technology that is affecting this process is digital lending, such as lending via mobile phone apps. Today, everything is expected to be done electronically and a decision given in minutes. So, overall, the decision management process has not changed – banks still need to make decisions based on profitability and onboarding the right customers. But how they do it and the speed at which they do it have certainly changed. So, how can we make digital decisions quickly? Firstly, we need to get all data onto a digital format. Secondly, we empower banks, mostly smaller institutions and FinTechs, using non-traditional data, to make those decisions.

R&C: In what ways does the availability of non-traditional data affect the process for credit decisions?

Akhadov: Organisations are awash in data. The challenge is creating meaningful connections between data scattered throughout an organisation and derive insights to support improvements in operational efficiencies, decision making, customer experience and profitability. Adding non-traditional data will add to the complexity from a compliance and execution perspective, while the overall credit decisioning process, if optimised already, will remain the same. There are different ingredients required for success, such as people, process, technology platforms and governance. Technology platforms are your tools. Analytics is the primary enabler to derive meaning from data and turn it into actionable insights. The tools are only as powerful as the people and process a company has developed. That said, an organisation could have a great data scientists team with well-thought-out processes, but if its tools are terrible, it may lead to failure due to various issues like poor data quality, challenges in deployment and proper model maintenance in production.

Siddiqi: Typically, when you apply for a loan on your mobile phone you agree to give the bank or lender access to the data on your phone. Online, when you agree to give someone access to your account, it is basically the same thing. So if you are an existing customer of the bank, they will look electronically at your internal history, savings and chequing account, existing loan information and credit score, and make a decision. In fact, there is software out there that enables banks to do that in a millisecond for an existing customer, with all that information readily available. When banks have a customer whom they do not know, there is no existing data, so they read the data on their mobile phone, and if they have access to that individual's social media profile, they may also use this data as part of the decision-making process.

"Over the past 10 years we have observed dramatic changes in consumer needs, behaviour and expectations. Consumers are now seeking easy access, choice, better control and speed."

Abdullo Akhadov, SAS Institute Inc.

Filipenkov: Non-traditional data allows organisations to make credit decisions faster and more accurately. The more data you have, the better equipped you are to make an accurate decision, which means a lower default rate, more customers and less churn. On the other hand, the use of non-traditional data may require specific knowledge and tools. A substantial portion of non-traditional data may be unstructured, such as web search requests or social media posts. So, one has to use text mining tools to make decisions based on this data. Similarly, one has to approach image recognition process and biometric data. Even if the data is structured – such as global positioning system (GPS) data – a data scientist has to introduce new features based on this data and use various ensemble techniques to combine the models based on traditional and non-traditional data.

R&C: When talking about non-traditional data, most people usually think of social media data. Are there other, better types of non-traditional data? What should bankers use?

Filipenkov: Bankers use a variety of different data sources, which may differ significantly based on the country where they operate and its data regulation legislation. A broad list of external data sources may include credit bureau data, various data provided by government, such as tax and social security data, energy and telecom payments, social media data, and data that comes from mobile apps like geolocation and browsing history. Some banks are introducing game apps and quizzes to understand the personality of their clients. Bankers should use all existing information allowed by local data protection and privacy legislation. Also, this would depend on

"Bankers use a variety of different data sources, which may differ significantly based on the country where they operate and its data regulation legislation."

Nikolay Filipenkov,
SAS Institute Inc.

the financial institution's appetite for model risk. The more models you use, the harder it is to manage model risk.

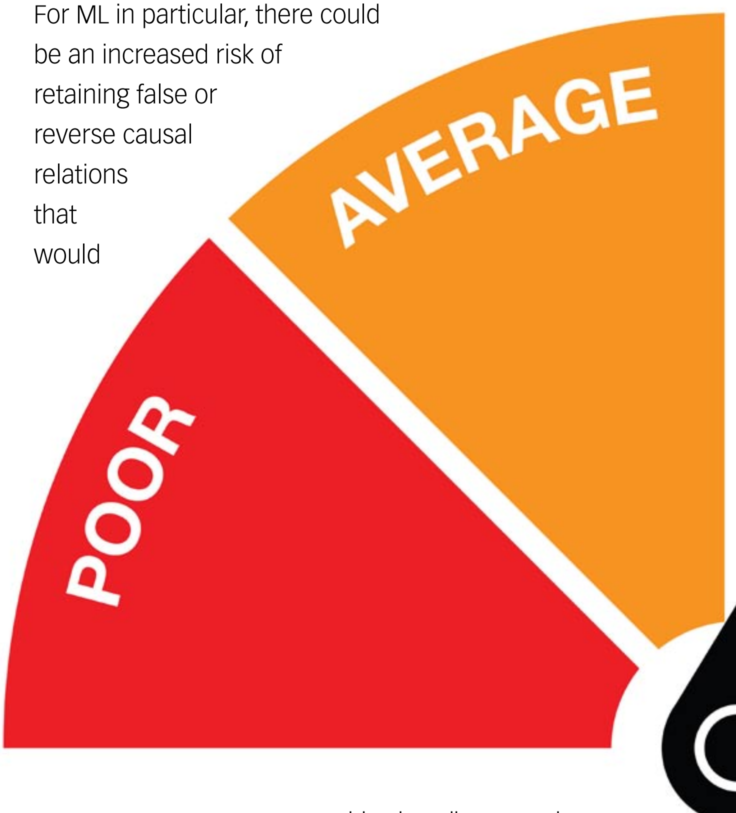
Siddiqi: When we talk about non-traditional data, also known as alternative data, people normally think about social media data. However, social media data is the worst kind of non-traditional data to use for lending. There are much better types. Some banks are starting to use utilities payment data, as well as mobile phone and telephone bill payment data, to make decisions on smaller loan amounts. Someone who has paid their phone, water and electricity bills consistently for the last 10 years and never missed a payment is probably a good customer. Such data is directly relevant to credit and creditworthiness – not necessarily for a $500,000 mortgage, but certainly for small loans. This is the type of data I would normally encourage banks to use in order to make decisions on people they do not know. Social media data is problematic as it is prone to fraud, coming as it does entirely from the customer. Likes and follows on Facebook and Twitter can be linked to creditworthiness. So, if you 'like' and follow Bloomberg, Reuters, Forbes, Fortune, The Financial Times, The New York Times, Wall Street Journal and The Economist, your scores will tend to be higher as they are linked to social strata or income level.

Akhadov: So, what can be considered as alternative data? Data from telecoms has information on customer behaviour and a lot of telcos around the world are trying to monetise it, either by moving directly into the lending business or selling data and credit scores to banks. Mobile data is also considered to be alternative data: how you use your phone, data, apps and calls. There are various FinTech start-ups collecting this type of data and selling scores to banks. In addition, utilities, rental data, social media – LinkedIn profiles and groups, as well as Facebook groups and likes – browsing data, social networks, such as addresses, friends and colleagues, psychometrics, Internet of Things (IoT) information, transactional data from e-commerce and transactional payments data are being used. Some interesting points can be found in various studies done by Experian, the Fair Isaac Corporation and TransUnion, which show that 66 percent of lenders using alternative data were able to onboard more creditworthy customers, 56 percent of lenders using alternative data were able to expand into new markets, nearly two in three lenders saw tangible benefits from alternative data in a first year, and the use of alternative data can lift approvals 60 percent within the same level of risk. What is really important is that consumers are ready to share their data with lenders if it will increase the probability of approval and give them a better interest rate. The predictive power of data varies from country to country and by consumer segments, so there is no one size fits all solution for banks.

R&C: What are the main challenges in the use of non-traditional data and machine learning (ML) techniques in credit decisioning?

Siddiqi: The most common things I have seen bankers use are random forests, gradient boosting, neural networks and support vector machines, which are hard to explain. They are what is known as a black box – you get an output but you do not know where that output came from and have no idea as to what sorts of variables or data is used inside algorithms to make a decision. They also tend to overfit. On the one hand, you have a black box method which is prone to overfitting, and then if you add dubious data from the internet or social media, then you have a bigger problem. For example, I have seen the data set of one online lender that had been web-scraping social media profiles, including social media likes and follows. This may indicate a left-wing or right-wing bias, depending on what you are reading. It may also indicate income level. However, in countries such as Canada, the US, the UK or Australia – places where there is a large immigrant population – you also have people who will follow ethnic media. An Indian immigrant may follow an Indian newspaper or television station. A Chinese immigrant may follow Chinese equivalents, and so on. This may cause compliance and legal issues as decisions may be made based on race, religion and ethnicity.

Akhadov: Before bringing alternative data-based decisioning into production, an organisation needs to understand if selected data from alternative data source can produce stable outputs. Is data amount and quality sufficient to build a model, considering that alternative data will require high granularity in customer segmentation for better differentiation? Is data either unbiased or is an organisation treating these biases, such as gender or race, before model development data is adjusted and during model development special penalty functions are introduced? An organisation also needs to know if data can be attributed to a specific customer and legal consent can be obtained. While a recent report by the Financial Stability Board (FSB) concluded that ML applications can be promising if their specific risks are properly managed, there are many challenges associated with artificial intelligence (AI) and ML models. These include the possibility that the use of complex, black box algorithms could result in a lack of transparency. A major assumption of any model calibration is that the sample data used is representative of the wider population. For ML in particular, there could be an increased risk of retaining false or reverse causal relations that would



otherwise be easily detected and removed by human intelligence. In addition, the instability of the credit model that may result from ML, due to continuous learning, could make it difficult to assess, challenge, validate and supervise the models or the algorithms used to calibrate them. The supervision or auditing of ML models may be difficult or even impossible, as it could be unclear which model is currently in use, how it was calibrated, how it may change over time when new data is added, or exactly how certain variables are used by the model, increasing operational and model risk. The analysis of data points through an automated process and without human judgement may indirectly lead to discrimination by considering sensitive characteristics, such as race, sex and ethnicity. If all you have is traditional homogenous data, then most probably a traditional model like linear regression will perform better and be more stable, also requiring less effort from a development and implementation perspective. Moreover, if you have a lot of detailed alternative data, where correlations are complex and deeply hidden, this is where ML models may add value.

Filipenkov: There are many challenges in the use of non-traditional data and ML techniques in credit decisioning. One of them is the data protection regulations in many countries. These may significantly limit the availability and applicability of many non-traditional data sources. Even if regulations allow the use of non-traditional data, the second challenge is the quality of this data. A lot of data has to be cleaned and pre-processed in order for it to be used in the decision-making process. Once the data is pre-processed and ready for use by data scientists, the third challenge is the skillset and tools required to build ML algorithms based on this data. Many non-traditional data sources such as text are different from the data sources that data scientists in banks are accustomed to working with. Finally, when a data scientist successfully builds a model, the challenge is to operationalise this model and use it in production. With many decisions based on ML algorithms needing to be made in real-time, infrastructure is becoming more important. And even when a model is implemented successfully, there may be more regulations that may limit the use of such a model.

R&C: How can organisations make sure that the use of alternative data and ML techniques are compliant with regulatory expectation and ethics?

Akhadov: Always ensure there is user consent for use of each specific alternative data source and item and variable, and motivate users to share. Do proper due diligence of data quality and data biases related to alternative data sources. Ensure established corporate policies are in place which take regulatory compliance and ethics into account, especially when it comes to predictive models and data used for model development. Implement control functions overseeing the data acquisition and model development process from a regulatory compliance and ethics perspective. Establish ML model performance monitoring procedures, and automate and allocate resources to run this process daily. Finally, employ strong data scientists and use best in class analytical platforms, which together should enable an organisation to build explainable and unbiased ML models.

Filipenkov: Regulatory expectations vary from country to country and change over time. Each organisation decides how it uses alternative data and ML techniques based on its risk appetite, code of ethics and position in the market. If there is enough traditional data or regulation is strict, an organisation may not be incorporating alternative data into the lending process. Conversely, if traditional data is limited – a lack of data from public authorities or credit bureaus – and regulation lax as to the use of non-traditional data sources, an organisation that uses non-traditional data sources may receive a significant competitive advantage and rapidly grow its market share and profitability. Some banks clearly state that they rely on ML to gain a competitive advantage.

> **"If you are going to use alternative data and ML, my suggestion is to look at the data you are using. Make sure it is clean and contains nothing dubious."**
>
> *Naeem Siddiqi, SAS Institute Inc.*

Siddiqi: If you are going to use alternative data and ML, my suggestion is to look at the data you are using. Make sure it is clean and contains nothing dubious. Also, ensure that all variables have some causal links to credit risk and creditworthiness. It is important to do your homework and due diligence on traditional and non-traditional data and then feed it into a ML technique. Then you build the model. There are different algorithms that attempt to explain such ML models. Once you have the black box model producing an output, what I suggest is to analyse the output to ensure it is stable. There are statistical tests for stability which back-test the model, validate it and benchmark it. What this does is provide a sense of comfort that the model is stable, produces reasonable results and is similar to the open box models that have been used in the past. This provides a sense of comfort that the model will perform as expected, with no ethical issues.

R&C: In what ways would you suggest that traditional banks change to deal with the current disruptive players?

Siddiqi: Traditional banks have to deal with disruption. There are a lot of destructive players, with many FinTechs nipping at their heels. We are seeing an increasing number of FinTechs in the microfinance industry, as well as the unsecured, higher-risk loans industry. So one thing traditional banks have to do is figure out what business they want to be in. FinTechs may be lending $1000 or $2000 based on social media data or mobile phone data, but do banks really want to compete in that space? Are they even in a position to compete with them and take on the associated risks? One thing traditional banks have to do is invest in their infrastructure and architecture, because many FinTechs are using 2018 technology to lend in 2018, whereas banks are literally using 1995 technology. This disparity is what causes most of the problems for banks. It is not mathematics; they can buy ML software very easily. The problem is taking ML models and deploying them quickly. In most banks, it still takes six to nine months to deploy a model. This is an infrastructure issue.

Filipenkov: Current disruptive players are entering the financial services market and traditional banks have started to react. Many are creating marketplaces and involving other companies in their ecosystem to get more data on the behaviour of their customers. Some banks have already done so, with some partnering with technology companies. There are also financial institutions claiming that they are transforming themselves into IT companies in order to compete with disruptive players. Banks have to change and use all available data sources and ML techniques to be competitive.

Akhadov: Overall approaches will vary depending on the market, internal organisation and politics – being able to transform into an agile organisation, support innovation culture and be ready to invest in research, as well as utilise seed investment, for example. Banks are now exercising various approaches, which all have pros and cons. These include in-house innovation labs, public innovation labs, partnerships with external innovation labs, partnerships with vendors and consultants, partnerships with venture capitalists, partnerships with FinTechs, investment into FinTechs, acquisition of FinTechs and innovation alliances with other banks. The European Banking Authority (EBA) published two papers in July 2018 which examined the impact of FinTech on incumbent credit institutions. EBA findings include factors that could potentially threaten the sustainability of incumbents' business models, such as digitalisation strategies that incumbent institutions pursue to keep up with the pace of a fast-changing environment, legacy information and communications technology (ICT) systems, operational capacity to implement the necessary changes, the ability to attract and retain staff and increasing competition from peers and FinTech firms. The EBA highlighted that many banks are rethinking their business models and are trying to transform their processes and technology platforms to become more agile, and quickly adapt to changing circumstances in the market. One can conclude that in particular, there is a need to address legacy risk management platforms that do not allow them to easily and quickly test innovative use cases based on advanced analytics and non-traditional data. R&C