Artificial Intelligent Credit Risk Prediction: An Empirical Study of Analytical Artificial Intelligence Tools for Credit Risk Prediction in a Digital Era

Diederick van Thiel
AdviceRobo
Tilburg University

W. Fred van Raaij
Tilburg University

Millennials service expectations drive transformation from traditional lending into digital lending. The CAGR for digital lending is 53% until 2025. Therefore, in this growing information age new methods for credit risk scoring could form the central pillar for the continuity of a financial institution. This paper contains the first research into AI application in individual risk assessment across two advanced lending markets. The research has been performed on 133.152 mortgage and credit card customers of 3 European lenders during the period January 2016 – July 2017. As candidate models, we chose neural nets and random forests. The research describes three experiments that develop the artificial intelligent probability of default models. In all experiments AI models performed better than the traditional models. Scalable automated credit risk solutions can therefore build on AI in their risk scoring.

Keywords: Credit, Risk Scoring, Digital Lending, Lending Robotization, Big Data, Artificial Intelligence

INTRODUCTION

Context: The Playing Field a Decade After the Credit Crunch

Since the early existence of banks, during Italy's renaissance, proper risk management has always been the cornerstone of banks. According to Brown and Moles the global credit crunch, which began in 2006 with sub-prime mortgages in the United States, has highlighted the fundamental importance of the credit decision. (Brown, et.al., 2014) The credit crunch had a combination of drivers. Firstly, the financial health of credit bases decreased due to the intensive sales of sub-prime mortgages. Secondly, the risk was not adequately priced on an individual level, because risk management approaches did not work on such a level. Thirdly, financial innovation led to the product: 'asset backed securities'. (Mizen, 2008) These asset-backed securities led to the global spread of risks without fully understanding their location. When house prices swiftly dropped, a significant number of financially unhealthy borrowers were unable to pay their rent. Banks had few buffers by which to cope in terms of increasing defaults and the structured product of 'asset-backed securities'. This caused a global threat to the financial system. As the problems with sub-prime mortgages unfolded, unsound credit decisions came to light; how to manage credit risk had effectively been ignored or never learned. The huge loan losses sustained by banks and others caught up in the credit crunch when money lent was not paid back, underline the major impact of credit risk and – by implication – credit risk management on the financial health of their individual business and private customers. Credit risk is the risk of default on a debt that may arise from a borrower's failing to make required payments (Basel Committee, 2000). This al shows that poor lending decisions, such as over-crediting or mispricing, led and will continue to lead to significant losses and further threats to the global financial system.

Now, a decade later, we see that the causes of the 2008 credit crunch remain unsolved. On the contrary, globally the outstanding amount of credit doubled compared to the lending volume of 2008 and continuously more credit decisions are being taken. Currently, not only banks, but also tech-giants like Amazon and Alibaba, have rapidly entered the lending market. The strong growth of lending in online retail, developing markets and peer-2-peer lending has infected the quality of credit bases once again. Central banks have lowered interest rates to levels that disable the interest instrument as an economic downturn appears. An economic downturn threat leading to a new financial crisis is currently being caused by global uncertainties like the USA-China trade war, currency crises in developing countries, wars, climate change and other instability causing things. A decade after the credit crunch, the global financial system is again at high risk of collapse.

Another transforming development in credit is the changing demands from millennials for customer experience. Millennials drive a change in customer experience expectations. The digitalization as a result of this, transforms borrowers in data agents producing tons of behavioral data that might contain differentiating risk features. New analytical methods are required to apply this combination of structured and unstructured data. The global market for digitization of lending will grow at a CAGR of 53% to $ 83.6 billion in 2025 (Zion market research, 2018). Digitization allows lenders to more effectively target their customers with appropriately timed offers. Digital lending automates complex processes and reduces manual interferences owing to which its demand is increasing. In the coming years, there will be an increasing adoption of digital lending.

Customer experience and financial advice are ill-defined concepts, and lack well developed assessment methods and metrics (Van Thiel, et.al., 2017). The influence of self-directedness on financial decision-making increases because the Internet enables consumers to learn from the experiences of others and to gather product information. In their research van Thiel & van Raaij developed the DCX-model that reveals the factors and attributes that drive customer experience toward digital financial advice models. Driven by the digitalization of customer experience, consumers become data agents themselves. These data might become very useful to improve credit decisioning in the upcoming decade. New analytical technologies need to be adapted for the application of this behavioral data.

The purpose of this paper therefore is to provide a contribution to the improvement of individual credit decisioning. This paper contains the first research across the United Kingdom and The Netherlands on how to improve credit decisioning with advanced modelling techniques like random forests and neural networks. Two highly advanced European lending markets that were seriously impacted by the 2008 credit crunch selected to examine to what extent lenders can advance their credit decisions with individual risk assessments on artificial intelligence. The research has applied supervised learning and has been performed on 133.152 mortgage and credit card customers in prime, near prime and sub-prime lending segments of 3 European lenders across the UK and the Netherlands during the period January 2016 – July 2017. As candidate models, we chose neural nets and random forests, as they carry the benefit of being able to work with both structured and unstructured data.

Credit Risk Management

Credit risk can be defined as "the potential that a contractual party will fail to meet its obligations in accordance with the agreed terms" (Brown, et.al., 2014). As a result of transactions of various kinds, credit risk and credit risk management are key issues for most firms (Brown, et.al., 2014). The possibility that a contractual arrangement is not adhered to equates to the risk of non-performance. This has the capacity to hurt the objectives of a firm; when a strategic plan is drawn, and it does not happen. Money can be lost if the customer fails to pay, or if the financial institution in which money is deposited, goes bankrupt. Companies with whom the firm has placed orders may themselves become insolvent and fail to deliver on their promises. There are three characteristics to define this credit risk:

1. Exposure at default (to a party that may possibly default or suffer an adverse change in its ability to perform).
2. Probability of default. The likelihood that this party will default on its obligations (the default probability).
3. Loss severity or its inverse the recovery rate (that is, how much can be retrieved if a default takes place).

In this paper, we define the business issue as the prediction of non-performance (probability of default); also, the larger the first two elements, the greater the risk. On the other hand, the higher the amount that can be recovered, the lower the risk. Formally, we can therefore express the risk as:

Credit risk = Exposure at default * Probability of Default * (1- Recovery Rate)

While the credit decision is relatively straightforward in theory (a lender must decide whether to give credit or refuse credit to a potential client), in practice it involves experience, judgement and a range of analytical and evaluative techniques that are designed to determine the likelihood that money will be repaid or, equally, that the money will be lost (borrower unable to repay). Managing credit risk therefore is a complex multi-dimensional problem, and as a result, there are a number of different, often portfolio-based, approaches in use - some of which are quantitative, while others involve qualitative judgements. Whatever the method used, the key element is to understand the behavior and predict the likelihood of borrowers defaulting on their obligations (Brown, et.al., 2014).

To understand the behavior and to predict default, all methods follow the same process and risk management framework; namely, identification, evaluation and management. That is, the cause of the risk must be identified, the extent of the risk has to be evaluated and decisions have to be made as to how this risk is to be managed.

The first step in the credit management process is to identify the problem (Brown, et.al., 2014). In most cases, we look simply at the no-default/default probability variable. In some applications it might be more complex, since we may want to monitor and evaluate changes in credit quality, rather than simple non-performance only. Irrespective of how the initial problem is defined, the size of the problem is then evaluated. Knowledge based models (expert models), effect models and statistical models are applied here. However, these require data and/or information from the business environment (i.e., application information, payment history information and personal information). The different analytical approaches for this can be loosely grouped into: (1) knowledge models, which have a degree of subjectivity (i.e., the use of expert judgement by an analyst); (2) effect models, which combine some elements of subjectivity and systemic analysis (a ratio analysis would fall into this category); and, (3) statistical models, which can be considered a more systematic approach (such as, credit scoring models).

Model validation or, measuring the quality of the probability of default models, can be conducted in several ways (Stein, 2002). Model validation becomes increasingly important as artificial intelligent approaches with a black box character contain a serious risk to model risk. Model risk is loss resulting from using insufficiently accurate models to reach decisions (Derman, 1996). When assessing the quality of a PD model, Stein differentiates model predictive power and model calibration. Model power describes how well a model differentiates between non-defaulting (good) and defaulting (bad) customers. A common statistic for assessing model power is the ROC-curve. ROCs are constructed by scoring all credits and ordering the non-defaulters from worst-to-best on the x-axis, then plotting the percentage of defaults excluded at each level on the y-axis. Here, the y-axis is formed by associating every score on the x-axis with the cumulative percentage of defaults with a score equal to, or worse than, the score in the test data. In other words, the y-axis gives the percentage of defaults excluded as a function of the number of non-defaults excluded (Stein, 2002). A similar measure, a CAP (Cumulative Accuracy Profile) plot

(Sobehart, et.al., 2000), is constructed by plotting all test data from worst-to-best on the x-axis. Thus, a CAP plot provides information on the percentage of defaulters that are excluded from a sample (TP rate), given that we exclude all credits, good and bad, below a certain score.

CAP plots and ROC curves convey the same information in slightly different ways. This is because they are geared toward answering slightly different questions. CAP plots answer the question: *How much of an* entire portfolio *would a model have to exclude to avoid a specific percentage of defaulters?*

While, ROC curves use the same information to answer the question: *What percentage of* non-defaulters *would a model have to exclude to exclude a specific percentage of defaulters?*

The first question tends to be of more interest to businesspeople, while the second is somewhat more useful when analyzing error rates. Model calibration is transforming classifier scores into class membership probabilities (Walker, 1996). Calibration of credit model leads to cut off points in accepting new customers, limiting settings and credit pricing. In this research, we aim to test if artificial intelligent models have a better quality than traditional logistic regression models. Kaplan defines artificial intelligence as a systems ability to correctly interpret external data, to learn from such data and to use those learnings to achieve specific goals and tasks through flexible adaptation (Kaplan, et.al., 2019). Observation 1 to be tested in the context of this paper is therefore:

Observation 1: *Artificial intelligent models, like random forests and neural networks can qualify to improve credit decisioning in different asset classes like mortgage loans and credit card loans.*

Regulation Drives Credit Risk Innovation

In response to the credit crunch, BASEL III, a global risk framework, was developed to increase banks' liquidity and decrease their leverage (Basel Committee on Bank Supervision, 2010). Basel III is a global, voluntary regulatory framework of banks' capital adequacy, stress testing and market liquidity risk. The original Basel III rule from 2010 required banks to fund themselves with 4.5% common equity (up from 2% in Basel II) of risk-weighted assets (RWAs). Since 2015, a minimum Common Equity Tier 1 (CET1) ratio of 4.5% must be maintained by the bank and increased with an additional buffer of 1.5%. This brings the minimum Tier 1 capital on 6% of common equity. Looking forward, stricter regulations, which will apply to Amazon and Alibaba, among other new entrants, on capital buffers is to be expected after 2022 in Basel IV. However, driven by new entrants, we also expect simpler and more standardized models for credit risk in Basel IV.

The transformation to Basel IV has already started through the transformation of accounting principles of financial instruments to be introduced in 2022. As another response to increased risk levels of lenders, namely, the International Accounting Standards Board (IASB, 2010), promulgated stricter accountancy rules under the IFRS-9. Also, the Financial Accounting Standards Board (FASB) published CECL standards with comparable requirements for US credit institutions in June 2016. Both IFRS-9 and CECL contain stricter guidelines for impairment. Therefore, lenders are challenged to transform from historical portfolio-based credit risk buffering to individual and forward-looking credit risk buffering. In IFRS-9 the allowance will be based on expected losses from individual defaults over the following 12 months, unless there is a significant increase in credit risk. If there is a significant increase, the allowance will be measured as the present value of all individual credit losses projected for the instrument over its full lifetime. If the credit risk recovers, the allowance can once again be limited to the projected credit losses over the 12 months. Credit risk management transforms from application and historically-driven to behavioral, predictive and even prescriptive-driven. Innovation in credit risk management will, under pressure of regulation, focus on risk prediction and risk prevention per individual to structurally lower defaults and increase the financial health of customers. Therefore, 21st century advanced credit risk management will have to merge statistics, accountancy and financial management with behavioral and computer science to continually monitor the financial behavior of consumers, thus prevent risk.

There are some issues to overcome, however. As, under BASEL II already signaled, one of the big issues defined for proper credit risk management is the poor availability of robust data to quantify banks' risk (Basel Committee on Bank Supervision, 2000). Under IFRS-9 and new Basel regimes coming up, data availability and quality will become more important and banks are lagging external data adaptation, such as FinTechs and other tech-giants. Another issue to overcome is how to effectively find ways that benefit from the increasing amount of data while minimizing the risk of information overload. Thirdly, the increasing focus on privacy in our digital age will lead to stricter regulations, such as the General Data Protection Regulation (GDPR). Consumers need to be able to view, update or delete their personal data with banks, and lenders must give specific consent for all applications of their personal data. Finally, in Europe, the Payment Service Directive 2 (PSD-2) is currently being implemented; a game changer. PSD-2 obliges banks whose customers empower a third-party service provider to access their personal data and to provide transaction data to such third-party service providers. The data-explosion that will be caused by the PSD-2 will strongly impact risk management, and it will also raise issues for risk managers around digitally-based trust, identification and authentication.

As traditional credit risk management is driven by historical data, portfolio management and logistic modeling, such statistical models are unable to cope with these transformations being enforced through legislation, and they are also unable to cope with unstructured data and can therefore not benefit from the behavioral data explosion in delivering advanced risk management solutions, such as continuous individual monitoring, predictive and prescriptive services that are expected to drive customer experience. The purpose of this paper is to assess the opportunity of artificial intelligence technologies, like random forests and neural networks, driven by behavioral data as a solution for the increased global credit risk. Here, experiments have been done to test the benefit of statistical artificial intelligence in credit risk for the probability of default in consumer lending. In the experiments supervised learning was applied to classify good and bad payers with the AI-models.

The Digital Consumer: Big Data, Artificial Risk Intelligence and Risk Robotization

Driven by the global digitization of lifestyles, the world is currently experiencing a behavioral data explosion (Van Thiel, et.al., 2017). Click streams, transaction histories, social media, mobile behavior, psychographic surveys and sensors provide huge volumes of behavioral data. New credit decisioning applications are being developed. Many households in developing countries for example lack formal financial histories, making it difficult for banks to extend loans, and for potential borrowers to receive them. However, many of these households have mobile phones, which generate rich data about behavior. Björkegren and Grissen show that behavioral signatures in mobile phone data predict loan default, using call records matched to loan outcomes. (Björkegren, et.al., 2018) Van Thiel & Van Raaij show that psychographic features that provide insight in attitudes, lifestyles and values predict customer engagement. (Van Thiel, et.al., 2017) Van Thiel further researched the application of psychographic data on credit decisioning within AdviceRobo. And Zhang et al. show, in order to reduce the serious problem of information asymmetry between both sides of P2P loans, the use of social information to describe the behavior characteristics of the borrowers. (Zhang, et.al., 2016) A person's social behavior and language can reflect the characteristics of their behavior, which can be used as credit data. On the internet, the behavior and language of users can be obtained from social media. An increasingly number of data sources with potentially more classifying and predictive features will follow in the upcoming years.

Every day, 2.5 quintillion bytes of data are created, and 90% of data in the world today, were already produced within the past years (IBM, 2016). Our capability for data generation has never been so powerful and enormous since the invention of Information Technology in the early 19th century (Wu, et.al., 2014). The most fundamental challenge for Big Data applications is to explore large volumes of data and extract useful information or knowledge for future actions (Rajaraman, et.al., 2011). In many situations, knowledge extraction must be highly efficient and close to real-time, because storing all observed data is infeasible.

Big data means more than simply larger storage requirements or collecting data from social media platforms with millions of participants (Flood, et.al., 2016). ‘Bigness’ is a symptom of scalability issues in one or more dimensions – namely, the three Vs: volume, velocity and variety (IBM, 2016).

- Volume – Roughly speaking, this is the simple size in bytes of a dataset, which can place a strain on storage and computational resources (Flood, et.al., 2016). ‘Big’ means that organizations must increasingly deal with a peta-byte scale of data collection through click streams, transaction histories, sensors and elsewhere.

- Velocity – The rate at which data arrive, which can strain network bandwidth and stream analytics (O'Hara, 2015). Organizations must increasingly apply the data fast for supporting their applications as f.e. fraud detection.
- Variety – The diversity of schemas, or formal structures, for data arriving from different sources, which can strain data-integration processes (Halevy, et.al., 2006). Data from different sources does not fit neatly into existing processing tools.

So, a dataset is too 'big' when it becomes computationally infeasible to process the dataset using traditional tools (MongoDB, 2016); new tools are required to apply the exploding volume of behavioral data. As most of this new data is unstructured, it requires new analytical models that can cope with both structured and unstructured data. New analytical techniques rely on mature commercial technologies of relational database management systems (DBMS); data warehousing; extraction, transaction and load (ETL); online analytical processing (OLAP); and, business process management (BPM) (Chaudhuri, et.al., 2011). Since the late 1980s, various data mining algorithms have been developed by researchers from artificial intelligence, algorithm and database communities. Most of these popular data mining algorithms have been incorporated in commercial and open source data mining systems (Witten *et al.*, 2011). Other advances, such as, neural networks for classification/prediction, clustering and genetic algorithms for optimization and machine learning have all contributed to the success of data mining across different applications (Chen, et.al., 2012). These scalable intelligent automated continuous, often platform, applications are considered the first risk robots. Assessing credit risk on the behavioral data of an individual might be more scalable than regression models that are very situation-specific; hence, the second observation to research is defined as:

Observation 2: *Artificial intelligent models, like random forests and neural networks can qualify to improve credit decisioning by having the ability to apply both structured and unstructured data.*

As consumer behavior becomes increasingly digital, generating an increasing volume of behavioral data, consumer lending will see further growth. Here, other elements like digital privacy, identification and authentication will have to be monitored prudently. Credit risk management will stay the most important element of post-credit crisis lending, but must re-invent itself accordingly. It will have to change from a historically portfolio-focused monitoring function to a pro-active predictive and prescriptive service for individual customers. As access to good data is considered one of the issues for proper risk management, data architecture and data cleaning will take priority. But with all pressure from society (privacy, digital trust), regulation (capital ratios and avoidance of individual risk) and shareholders (cost/income and capital ratios) - scalable 'risk robots' will likely standardize these highly complex forward-looking activities in coming years. Across many geographies, an increasing number of financial service providers are currently operating or considering utilizing, the use of robo-advisors - online platforms that provide advice using complex computer algorithms (Bradbury, 2014). These robo-advisors make use of the increasing amount of behavioral data and apply algorithms that match consumers or small businesses with financial products or portfolios (Van Thiel, et.al., 2017). The purpose of this paper is to test the impact on risk management of artificial intelligent techniques that will drive automated risk management for advice-robot solutions. Research has been performed to assess the extent to which the application of neural networks, random forest and support vector machines, results in better default predictions in a digital and heavily regulated global market. The research describes three experiments conducted across the United Kingdom and the Netherlands, which develop advanced probability of default models and compare the model quality with the quality of the traditionally applied PD-models. Butaru *et al* performed similar research on the data of US credit card lenders. (Butaru, et.al., 2014) The difference with this research is that we focus on different credit products over different geographies, while Butaru examined only the US. Also, Khandani *et al.*, performed research on artificial intelligence on risk prediction. (Khandani, et.al., 2010) The difference with Khandani and colleagues, is that they focused on one bank, while this research incorporates multiple banks. This all leads to our final observation:

Observation 3: *Artificial intelligence models predicting default risk can be applied across different geographies and product groups without having to customize them.*

METHOD

Empirical Design and Modeling Approach

Our dependent variable is the delinquency (default) status. For the purposes of this study, we define delinquency as a mortgage or credit card account greater than or equal to 90 days past being due (Basel Committee on Banking Supervision, 2016). We assume that we are solving a two-class classification problem; the learning algorithm takes the training dataset, consisting of pairs ($\mathbf{x}$, y), where $\mathbf{x} \in X$ is the feature or attribute vector (and can include categorical- as well as real-valued variables), and $y \in \{0,1\}$ as input. The output of the learning algorithm maps X to $y \in \{0,1\}$ (or possibly, in the case of logistic regression, to [0, 1] where the output represents Pr ($y = 1$)). To compare the quality metrics of the models and to standardize for robo-risk intelligence, banks participating in the experiments delivered the exact same dataset they themselves apply in their traditional logistic regression risk models. The mortgage data sets contained a three-year transaction history. The thin file credit card data set contained a 1-year transaction history. Because, bank datasets differentiated, to be able to draw learnings for an automated robo-solution across geographies, banks, customer - and product segments, we use the Azure Machine Learning Studio (see https://studio.azureml.net for more information) to run the same models on the different data sets.

Data Preparation

The first step is to collect and prepare the data. To avoid data-compliancy and privacy issues, participating banks shared anonymized customer data. All datasets collectively form a sample of 133,152 customers. The two samples of Dutch banks for mortgage default prediction are sized: 55.812 and 47.346. The sample for thin-file credit scoring of a British credit card issuer is 6,994, thus substantially smaller.

The data is prepared for the machine learning models using complete and coherent meaningful features. Also, assessments are conducted on data definitions, the data sources and banks' policy definitions of delinquencies. To compare outcomes with traditional logistic regression modeling approaches, sources applied for default predictions in these experiments are internal bank data only. However, in the UK experiment on credit scoring thin-file customers, external credit bureau data are also applied in the logistic regression approach. Here, we cooperated in the UK with credit bureaus Experian and CallCredit.

Having received the anonymized datasets, data-cleaning took place by deleting and repairing missing values. On average 0.26% of data was missing and 10.67% qualified as outliers. After cleaning the data, feature development was performed on all datasets. In feature preparation, we looked at (1) null-values; (2) whether a feature has a discrete or continuous character; and, at (3) meaningful ratios, like income to loan to be designed as new features. Discrete features were made binary, and to finally check the feature quality, statistical analyses per feature were performed, such as, f.e. calculating the maximum and minimum value, the mean, median and standard deviation. In addition, the sample data was partitioned following the hold-out method into a training (70%)/validation (30%) sample.

Model Development

After preparing the data, the candidate machine learning models were trained 50 times with a different sample of the training data. As candidate models, we chose random forests and neural nets since they are the most popular supervised learning methods that are able to work with both structured and unstructured data in credit risk.

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random decision forests correct for decision trees' habit of overfitting to their training set and show less variance by having more trees. Note that one major problem with decision trees is their high variance. Often a small change in the data can result in a very different series of splits, making interpretations somewhat precarious. The major reason for this instability is the hierarchical nature of the process: the effect of an error in the top split is propagated down to all splits below. One can alleviate this to some degree by using a more stable split criterion, but inherent instability is not removed. It is the price to be paid for estimating a simple, tree-based structure from the data (Hastie, et.al., 2009).

The random forest method combines two important ideas to improve the performance of decision trees, which are the base learners. The first idea is bagging, or bootstrap aggregation. Instead of learning a single decision tree, bagging resamples the training dataset with replacement T times, and learns a new decision tree model on each of these bootstrapped sample training sets. The classification model then allows all T decision trees to vote on the classification, using a majority vote to decide on the predicted class. The key benefit of bagging is that it greatly reduces the variance of decision trees, and typically leads to significant improvements in out-of-sample classification performance. The second key idea of random forests is to further reduce correlation among each of the induced trees by artificially restricting the set of features considered for each recursive split. When learning each tree, as each recursive split is considered, the random forest learner randomly selects a subset of the features (for classification tasks, typically the square root of the total number of features), and only considers those features. Random forests have been enormously empirically successful on many out-of-sample classification benchmarks in the last decade and are considered among the best 'out of the box' learning algorithms available today for general tasks (Caruana, et.al., 2006; Criminisi, et.al., 2012).

An artificial neural network is a network of simple elements called artificial neurons, which receive input, change their internal state according to that input, and produce output depending on the input and activation. The main advantages of using Artificial Neural Networks include the handling of large amount of data sets; the ability to implicitly detect complex nonlinear relationships between dependent and independent variables and the ability to detect all possible interactions between predictor variables. Major limitation for credit scoring is the black box character of the neural network as regulators demand lenders to be able to explain the reasons for accepting or rejecting new applicants. For this reason, we used the neural networks to understand their impact, but focused on the random forest models in reporting to the lenders in the experiments.

Measuring Performance

The goal of our delinquency prediction models is to classify mortgage and credit card accounts into two categories: accounts that become 90 days or more past due within the next n quarters ('bad' accounts), and accounts that do not ('good' accounts). Therefore, our measure of performance should reflect the accuracy with which our model classifies the accounts into these two categories.

One common way to measure performance of such binary classification models is the AUROC. The AUROC, or Area Under the ROC-Curve, is a score between 0 and 1 that shows the predictive power of a model by calculating the mean of the precision and recall. Precision is defined as the number of correctly predicted delinquent accounts (true positives) divided by the predicted number of delinquent accounts (true positives + false positives), while recall is defined as the number of correctly predicted delinquent accounts (true positives) divided by the actual number of delinquent accounts (true positives + false negatives). Precision is meant to gauge the number of false positives (accounts predicted to be delinquent that stayed current), while recall gauges the number of false negatives (accounts predicted to stay current that went into default).

Although we primarily look at the AUROC to test our hypotheses, we know other statistics are also worth looking at when qualifying a model. Indeed, a widespread metric is the Gini-score. Gini is 2 times the AUROC – 1. Another metric is the F1-measure. The F-measure is defined as the harmonic mean of precision and recall and assigns higher values to methods that achieve a reasonable balance between precision and recall.

The other performance indicators to consider when selecting the champion prediction models are: (1) overall accuracy rate (bias between reference value and mean of the measurements); and, (2) stability of model (stable over time and different datasets).

The modeling approach can be summarized as follows:

(1) 50 variants within each modeling algorithm are tried and applied on the training sample. The modeling algorithms used are Neural Net and Random Forest.
(2) Each model is applied on the validation sample. The champion model within each modeling algorithm is identified based on the above performance indicators.
(3) The best performing champion models of the different experiments are analyzed on similarities in features, as well in type of model.

RESULTS

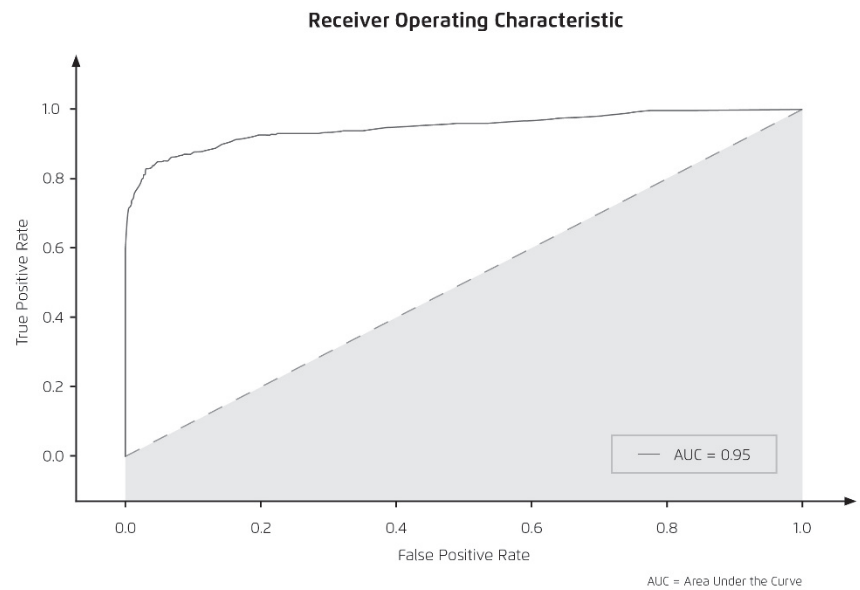
Experiment 1: Dutch Bank Insurance Company

The first experiment was held between January 2016 and August 2016 with a tier-2 Dutch bank insurance company. The bank services 47,347 mortgage customers and holds a mortgage portfolio of €10 billion. The bank's strategy focusses on improving customer experience and operational excellence. To improve their customer experience, they want to understand the opportunity that artificial intelligence provides for lowering default rates. To accomplish this, the bank stepped into this experiment to test the quality of their traditional logistic default prediction model against a machine learning champion model. Actual logistic regression area under the curve (AUROC) is 0.87 and actual defaults in 2016 were 0.9%.

The bank anonymized their customer data and securely shared 67 anonymized application and behavioral features per individual. After training different models, the champion model for their data proved to be a random forest.

The random forest champion model performs an AUROC of 0.95%. Compared to the traditional AUROC of 0.87%, machine learning shows an improvement in AUROC of 18.8%. For this bank, observation 1 "AI predicts default risk better than traditional logistic regression" seems true. The AUROC is represented in the Lorenz Curve shown in Figure 1.

FIGURE 1
LORENZ CURVE DUTCH MORTGAGE MODEL



The most predictive features for this bank's delinquency are exposed in table 1 below.

TABLE 1
DUTCH BANK INSURANCE COMPANY MOST PREDICTIVE FEATURE LIST

	Prediction			
Actual class		Good payers	Bad payers	Total
	Good payers	101463	937	102400
	Bad payers	75	195	270
	Total			102670

We get deeper insights into model performance by looking at the underlying statistics. The precision in this experiment is 0.99 good. It is calculated as the fraction of true positives (101.463) divided by the sum of true and false positives (102.670). The recall in this experiment is 0.99, which is also good. Recall is the fraction of true positives (101.463) over the total amount of relevant instances (102.400). The precision and recall of this random forest model are derived from the confusion matrix, shown in table 2.

TABLE 2
DUTCH BANK INSURANCE COMPANY CONFUSION MATRIX

	Good payers	Bad payers
AuROC*	0.95	
Accuracy	0.99	0.99
F1 score	0.99	0.28
Sensitivity	0.99	0.72
Specificity	0.72	0.99

*Area under the curve

The quality metrics of the applied random forest champion model are shown in table 3, below.

TABLE 3
QUALITY METRICS RANDOM FOREST MODEL

Dutch Bank Insurance Company
1. Loan to Value
2. Ratio mortgage / income
3. BKR (credit bureau) registration
4. Ratio credit / income
5. Age

In a statistical analysis of binary classification, the F_1 score (also F-score or F-measure) is a measure of a test's accuracy. It considers both the precision *p* and the recall *r* of the test to compute the score: *p* is the number of correct positive results divided by the number of all positive results returned by the classifier, and *r* is the number of correct positive results divided by the number of all relevant samples. The F_1 score is the mean of the precision and recall, where an F1 score reaches its best value at 1 (perfect precision and recall) and worst at 0. The F1 score of the champion model is with 0.99 also good.

So, also having looked into the other statistics of this experiment, we conclude that in this first experiment the random forest approach improved the predictive power of the credit decisioning with 8% calculated on the difference between the AUROC's.

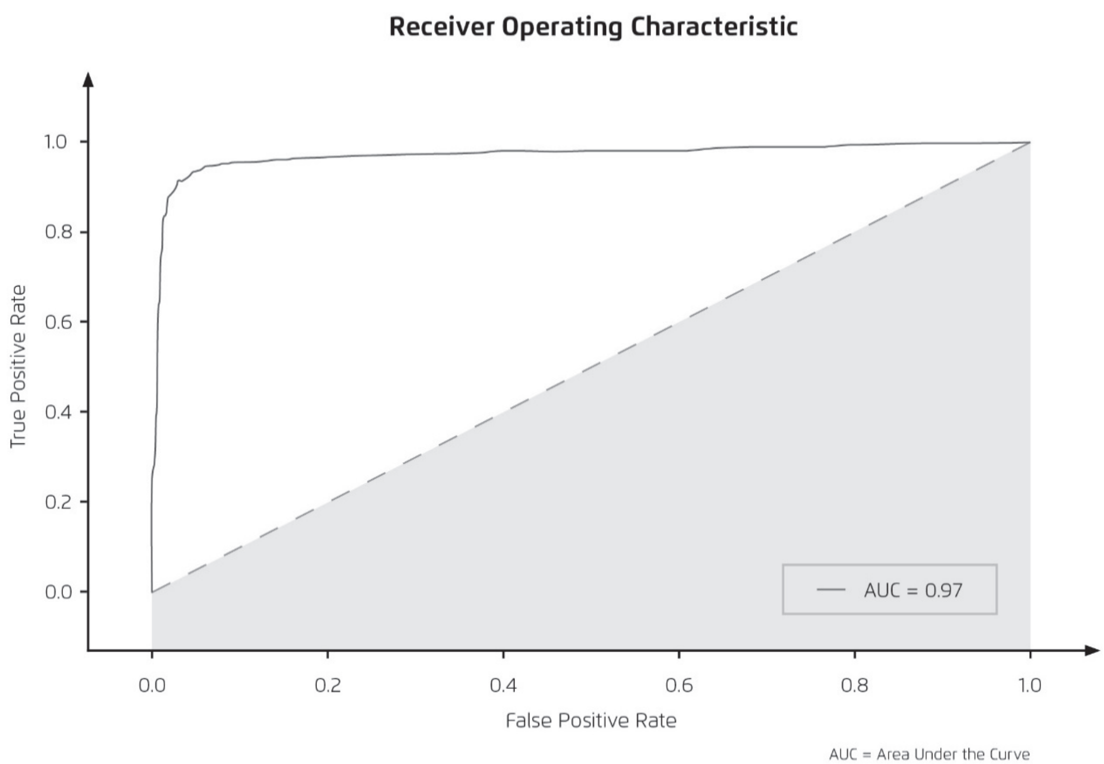
Experiment 2: Dutch Mortgage Bank

The second experiment covers the period January - July 2017 with a Dutch mortgage bank. The bank services 55,812 mortgage customers and holds a mortgage portfolio of €8.8 billion. Different from the other bank in this experiment, this has mortgage application data only. The bank's strategy focusses on improving customer engagement by being there at the most decisive moments in life. To improve their customer engagement, the bank wants to understand the opportunities that machine learning can provide in predicting default risk at their currently performing customer base (the data does not contain earlier arrears or delinquencies). The bank has an ambition to proactively support people, months before they experience mortgage payment problems. To accomplish this, the bank stepped into this experiment to test the quality of their traditional logistic default model against a machine learning champion model. The traditional logistic regression model gives a Gini-score of 0.8. As the AUROC = (Gini +1)/2 the AUROC is 0.9 and actual historical defaults in 2016 were 0.8%.

The bank applied the exact same method as we applied in the first experiment. They anonymized their customer data and securely shared 51 anonymized features per individual. For their pro-active servicing purpose, in this experiment we trained models predicting 6-month, 3-month and 1-month defaults. To be able to compare with the traditional logistic regression model, we focused on the 3-month (90 days) prediction model. After training different models, the champion model proved to be a random forest.

The machine learning champion model performs an AUROC of 0,97. Compared to the traditional AUROC of 0,8, machine learning shows an improvement in AUROC of 21,3%. The AUROC is represented in the Lorenz Curve shown in Figure 2.

FIGURE 2
LORENTZ CURVE DUTCH MORTGAGE MODEL



Also in this experiment we must assess the other metrics to fully prove the improvement mentioned before. The most predictive features for this bank's delinquency are shown in table 4.

TABLE 4
PREDICTIVE MODEL FEATURES

<table>
<tr><td></td><td colspan="3">Prediction</td></tr>
<tr><td rowspan="3">Actual class</td><td></td><td>Good payers</td><td>Bad payers</td></tr>
<tr><td>Good payers</td><td>985</td><td>18</td></tr>
<tr><td>Bad payers</td><td>65</td><td>466</td></tr>
</table>

In the second experiment, we also gain a deeper insight into the model's performance by looking at underlying statistics. The precision in this experiment is 0,94 and calculated as the fraction of true positives (985) divided by the sum of true and false positives (1,050). Recall is 0,98; the fraction of true positives (985) over the total amount of relevant instances (1,003). Both precision and recall also look good in this experiment.

The precision and recall of this random forest model are derived from the confusion matrix, as shown in table 5.

TABLE 5
DUTCH MORTGAGE BANK CONFUSION MATRIX

	Good payers
Accuracy	0.95
Precision	0.96
Recall	0.88
F1 Scores	0.92
ROC*	0.97
Gini Score	0.95
Matthews	0.88

*Receiver Operating Curve

The quality metrics of the applied random forest champion model are shown in table 6.

TABLE 6
QUALITY METRICS RANDOM FOREST MODEL

Dutch Mortgage Bank
1. Sum of interest
2. Payment of interest
3. Monthly payment
4. Net loan total
5. Net loan

The F_1 score in this experiment is 0.92, which implies an accurate model.

We can therefore conclude that also in the second experiment the random forest model performed better than the traditional logistic regression approach.

Experiment 3: British Credit Card Company

The third experiment covers the period of October 2016 and February 2017 with a British credit card issuer. The company services 5.4 million customers in the prime and near-prime customer segments. The credit card loan book is £1.8 billion. The company's strategy seeks the onboarding potential of thin-file customer segments. Thin-file consumer segments are segments with limited or no credit information. Therefore, thin-files have no access to credit. To access these customer segments, the company wants to understand the opportunity machine learning provides for onboarding thin-file consumers. To accomplish this, the company stepped into this experiment to test the quality of their logistic regression scorecard-model against a machine learning champion scorecard-model. Actual logistic regression Gini-score is 0.25 (thin-file customers) and actual impairment rate in 2016 was 8.8%.

The company gathered data on thin-file consumers by accepting 6,994 in three months' time; the company monitored thin-file customer behavior for 6 months, and the data gathered contains 20 features. Additionally, data from credit bureaus Experian and Call Credit were added. The company anonymized their customer data and securely shared the 901 features per individual. Because of the thin-file character of the customers, most features were empty and could not be used for modeling. After training different models, the champion model proved to be a random forest. The machine learning champion model performed an AUROC of 0.55 and a Gini of 0.32. Compared to the traditional Gini of 0.25, machine learning shows an improvement of 28%. Also, the application of machine learning on credit cards as well as the application in the UK seems to work.

Again, we must assess the more granular metrics to fully prove. The precision in this experiment is 0.79 and calculated as the fraction of true positives (978) divided by the sum of true and false positives (1,232). Recall is 0.94. Recall again is the fraction of true positives (978) over the total amount of relevant instances (1,041). Both precision and recall look good in this experiment. The precision and recall of this random forest model are derived from the confusion matrix, as shown in table 7.

TABLE 7
CREDIT CARD MULTI-BUREAU RANDOM FOREST MODEL

Neural Net	Prediction			
Actual class		Good payers	Bad payers	Total
	Good payers	3212	227	3439
	Bad payers	779	160	939
	Total			4378

The quality metrics of the applied random forest champion model are shown below in table 8.

TABLE 8
QUALITY METRICS CREDIT CARD RANDOM FOREST MODEL

	Good payers
Accuracy	0.77
Precision	0.41
Recall	0.17
F1 Scores	0.24
ROC*	0.55
Gini Score	0.32
Matthews	0.15

*Receiver Operating Curve

The F_1 score in this experiment is 0.24, which implies a less accurate model due to the small number of features available for this thin-file customer segment.

We nevertheless conclude that random forest performs better than logistic regression in all experiments. We can also see that AI models work across product groups and geographies. The shorter the payment cycle however, the better the models can be validated. Our observations nevertheless can be validated. It is not a complete validation, as the hypotheses were formed with the idea of standardization for robo-risk scoring in mind. The model features however differ to much across the experiments to be able to standardize yet. Further research on this needs to be conducted.

Results Summary and Observation Testing

The purpose of this paper is to assess the opportunity of analytical artificial intelligence technologies, namely random forests and neural networks, driven by behavioral data as a solution to improve individual risk decisioning. Here, three experiments have been conducted to test the benefit of AI credit risk models for probability of default in consumer lending. In all experiments, artificial intelligent models performed better than traditional models. The models of the British credit card company and the bank insurer, which could tap into payment data, perform better than mortgage-only data models. Payment of the interest and monthly payment are among the top predictive features with the mortgage-only bank. The bank insurance and credit card company looked more into credit score and loan to income ratio’s which they had access to. Looking at these most predictive features the models produced, high-level similarities can be uncovered across experiments. If banks have access to income or spending data, income or estimated income, or all of them this currently is an important feature for default prediction. We see more advanced lenders create more intelligent features by creating relations between income and loan and using social media data (bank insurance company). As explained before, in our analyses we primarily looked at the random forests as the neural networks black box character would not allow us to investigate the underlying differentiating features.

To validate the observations, results were made comparable between random forest and logistic approaches by applying the very same traditional structured dataset in the experiments and compare on clear risk metrics. However, the benefit of the more advanced artificial intelligence methods is that it can, on top of these traditional transaction data, also apply unstructured non-financial data groups to improve credit application scoring, risk monitoring and personalization strategies. Although the results support our contention that bank-specific calibrated models are likely to be better predictors of default as opposed to a single model applied to all banks, standardization of artificial intelligent models across banks and geographies seems to some extent possible. Further research has to be conducted in this area as it can bring an amazing cost reduction benefit to international banks if they can standardize their risk modelling across geographies and asset classes. Standardization might be started from more generic features. If, for example, a basic risk intelligence robot works with data like (total) loan versus income, the risk intelligence can be standardized for that part and both credit application and credit monitoring can be structured for that part across geographies. On top of that, modules with external scalable data groups like psychometric data, internet data, social media data and mobile phone data can make robotized risk intelligence even more sophisticated. Advanced artificial intelligence therefore seems to become the most powerful risk scoring approach in this era of robotization of risk management.

TABLE 9
OBSERVATION TESTING

	Experiment 1: Dutch Bank Insurance Company	Experiment 2: Dutch Mortgage bank	Experiment 3: Brittish Credit Card Company
Observation 1: Artificial intelligent models, like random forests and neural networks can qualify to improve credit decisioning in different asset classes like mortgage loans and credit card loans.	✓	✓	✓
Observation 2: Artificial intelligent models, like random forests and neural networks can qualify to improve credit decisioning by having the ability to apply both structured and unstructured data.	✓		
Observation 3: Artificial intelligence models predicting default risk can be applied across different geographies and product groups without having to customize them.	✓ ✗	✓ ✗	✓ ✗

DISCUSSION

Global consumer lending shows a compound annual growth rate (CAGR) of 4.8% up until 2020. However, specific segments like f.e. marketplace lending, show even higher growth. Market place lending shows a CAGR of 53.6%. As banks lend more money and new lenders pop-up, the risk of over-crediting and default increases. Better individual risk assessments, limit setting and pricing are required to reduce over-crediting.

Also, millennials drive a change in customer experience expectations. The digitalization as a result of this, transforms borrowers in data agents producing tons of behavioral data that might contain differentiating risk features. New analytical methods are required to apply this combination of structured and unstructured data. The global market for digitization of lending will grow at a CAGR of 53% to $ 83.6 billion in 2025. Digitization allows lenders to more effectively target their customers with appropriately timed offers. Digital lending automates complex processes and reduces manual interferences owing to which its demand is increasing. In the coming years, there will be an increasing adoption of digital lending.

In this study, we therefore employ a large dataset consisting of anonymized information from three banks in different asset classes like mortgages and credit cards across the United Kingdom and The Netherland, to test the added value of artificial intelligent risk models for predicting mortgage and credit card delinquency. The algorithms for mortgage lending have access to consumer transaction data with a three-year history and credit bureau data for credit card lending from January 2016 to July 2017. We find that random forests and neural nets outperform logistic regression in risk predictive power and have the ability to operate on both structured and unstructured data.

We also analyze and compare risk management practices across the banks and compare drivers of delinquency across institutions. We find that there is substantial homogeneity across banks in traditional risk features like payment of the interest, monthly payment, credit score and loan to income ratios. Nevertheless, the product mix of a lender strongly determines the availability of data and therefore, no single model is likely to easily capture the delinquency tendencies across all institutions, product groups and geographies yet. However, all of them currently focus on the relation between income and (total) lending amount so parts of the model could potentially be captured by a single cross product, cross geography model. The results also suggest that portfolio characteristics alone are insufficient in identifying drivers of delinquency, since the banks actively manage the portfolios. Even a nominally high-risk portfolio may have fewer volatile delinquencies because of successful and active risk management by the bank. The banks in the experiment are also substantial homogeneous in not applying external behavioral data yet. Only in one experiment we found the application of social media as part of the credit model.

Risk management practices on the other hand show heterogeneity across financial institutions which has systemic implications. Mortgages and credit card receivables form an important component of modern asset-backed securities. An unexpected macroeconomic shock may thus propagate itself through a greater delinquency rate of mortgages and credit cards issued by financial institutions who less actively manage their portfolio into the asset-backed securities market.

Our study provides an illustration of the potential benefits that advanced machine-learning techniques, and with that the use of unstructured data, can bring to consumers in terms of a faster and more predictive and prescriptive customer experience; to risk managers by transforming from expert driven modelling into digitalization of risk management with more advanced ways of artificial intelligent modelling and monitoring on more internal and external data and to shareholders by lowering delinquencies and regulators by better controlling systematic credit risk. All of them have a stake in avoiding unexpected losses and reducing the cost of consumer credit. Moreover, when aggregated across several financial institutions, the predictive analytics of machine-learning models provide a practical means for measuring systemic risk in one of the most important and vulnerable sectors of the economy. The AI-models show higher predictive power and the opportunity to scale risk models across product groups and geographies. Further research needs to be conducted at this scalability, but it will deliver a great benefit of high cost reductions and improved efficiencies in international risk management.

In this study, we develop random forest models for consumer credit delinquency, which is surprisingly accurate in forecasting credit events in three different experiments. Lenders can improve their credit acquisition and credit management strategies with more advanced machine learning. Traditional application data applied in machine learning models already improves scorecard performance. As consumers and lenders become more digital and mobile, adding behavioral data, both structured (f.e. payment data or credit card spending) and unstructured (f.e. search, sentiment, psychographics and mobile behavior) to these scorecards will further support sound onboarding and pricing strategies and will reduce mis-selling. Further research also needs to be conducted in this area of alternative data for risk scoring as it offers the benefit of breakthroughs in predictive model power, and therefore, gaining much better control on financial risks. Higher growth of the lending market is expected in developing countries. A significant 67% of the global population are thin-file (not credit rated) and therefore have no proper access to essential financial services; global citizens are unable to build up their lives and businesses. Globally, this covers 4.6 billion people. On the other hand, 89% of banks are unable to properly assess risk in information-poor environments. For companies willing to give these people access to whatever form of credit, new unstructured behavioral data in combination with machine learning offer good credit scoring solutions.

With the high growth of global consumer credit in a growing but unstable world economy, the need for better, individual and more effective risk assessments in lending bases becomes evident. Regulators enforce lenders with stricter capital requirements and IFRS-9/ CECL to do so. Traditional logistic early warning systems assess the portfolio of loans in a customer base on historical behavior. Artificial intelligence, on the other hand, offers lenders the opportunity to continually monitor individual risk development, based on behavioral structured and unstructured data. Also, the high predictive power of artificial intelligence offers opportunities for IFRS-9/CECL risk predictions and robotized solutions. Precondition obviously is the quality of the underlying data. Lenders who seriously want to improve their risk prediction should consider collecting behavioral data from all types of sources for improved feature development.

Finally, machine learning is often offered as a service by companies like Amazon, Microsoft and IBM. For specific machine learning services in credit risk management, FinTech has flooded the markets. Companies like f.e. AdviceRobo, Aire, EFL and Lendoo operate their platforms on an international scale. Serious cost reductions in the manual labor and legacy systems of risk management therefore becomes prevalent in supporting lenders' cost-to-income ratios. Risk robots are expected to bring high effectiveness to risk management in the upcoming 10-years. Their success will be highly dependent on finding and scaling the best predictive features within new unstructured data. Successful credit robots will reduce operational risk costs such as collections and fraud fighting on a huge scale. We plan to further explore the benefits and challenges of robotization & digitization of risk and marketing management in future research.

LIMITATIONS AND FUTURE RESEARCH

Limitations of this research center on the focus on two leading European credit markets. Similar research should be performed in other geographies, especially in developing countries. Secondly, the time frame of the experiments might bring bias. Although we could work with three-year historical data, the market changes rapidly. We therefore advice to repeat these experiments after a few years to understand the advancements in digitalization of risk management better. Another limitation is the application of structured data in order to make results comparable across models. Only in one experiment social media data was applied. The application of external and unstructured data is also something that might evolve over time. More research with the application of other data groups should be conducted to understand the impact of unstructured behavioral data on risk scorecards.

Also, further research needs to be conducted into the scalability of artificial intelligent risk models as the combined benefit of increased predictive power and higher international efficiency in risk management is present. In the context of mortgage and credit card portfolio risk management, there are account-specific costs and benefits associated with the classification decisions that our performance statistics fail to capture. In the management of existing lines of credit, the primary benefit of classifying bad accounts before they become delinquent is to save the lender the run-up that is likely to occur between the current time period and the time at which the borrower defaults. On the other hand, there are costs associated with incorrectly classifying accounts. For example, the bank may alienate customers and lose out on potential future business and profits on future purchases. This research does not calculate the financial impact per bank, but primarily focuses on the possibility of standardizing risk intelligence for the robotization of risk management.

ACKNOWLEDGMENTS

Manuel Martin Ortiz for performing analyses. Tomas Garcia for data and modeling support.

REFERENCES

Basel Committee. (2000). Principles for the management of credit risk. Basel Committee of Banking Supervision.

Björkegren, D., & Grissen, D. (2018). *Behavior revealed in mobile phone usage predicts loan repayment*. Available at SSRN 2611775.

Bradbury, T. (2014). Robo advice is coming: what it means, who will buy it-and why. *Professional Planner*, 69, 40.

Brown, K., & Moles, P. (2014). Credit risk management. *Credit Risk Management*, 16.

Butaru, F., Chen, Q., Clark, B., Das, S., Lo, A. W., & Siddique, A. (2016). Risk and risk management in the credit card industry. *Journal of Banking and Finance*, 72, 218-239.

Caruana, R., & Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms. *Proceedings of the 23rd international conference on Machine learning*, 161-168. ACM.

Chaudhury, K., Garg, A., Phukan, P., & Saraf, A. (2011). *U.S. Patent No. 7,986,843*. Washington, DC: U.S. Patent and Trademark Office.

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: from big data to big impact. *MIS Quarterly*, 1165-1188.

Criminisi, A., Shotton, J., & Konukoglu, E. (2012). Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. *Foundations and Trends® in Computer Graphics and Vision*, 7(2–3), 81-227.

Derman, E. (1996). Model Risk: What are the assumptions made in using models to value securities and what are the consequent risks? *RISK-LONDON-RISK MAGAZINE LIMITED*, 9, 34-38.

Flood, M. D., Jagadish, H. V., & Raschid, L. (2016). Big data challenges and opportunities in financial stability monitoring, *Banque de France, Financial Stability Review*, 20.

Halevy, A., Rajaraman, A., & Ordille, J. (2006). Data integration: the teenage years. *Proceedings of the 32nd international conference on Very large data bases* (pp. 9-16). VLDB Endowment.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *Unsupervised learning. The elements of statistical learning* (pp. 485-585). Springer, New York, NY.

IBM. (2012). Analytics: the real-world use of big data in financial services. IBM Institute of business value.

Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, 62(1), 15-25.

Khandani, A. E., Kim, A. J., & Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. *Journal of Banking and Finance*, 34(11), 2767-2787.

Microsoft. (2018). Retrieved from https://studio.azureml.net.

Mizen, P. (2008). The credit crunch of 2007-2008: a discussion of the background, market reactions, and policy responses, *Federal Reserve Bank of St. Louis Review*, 90.

Mongo, D.B. (2016). Retrieved from https://redmonk.com/sogrady/2016/07/01/mongodb-atlas/.

O'Hara, M. (2015). High frequency market microstructure. *Journal of Financial Economics*, 116(2), 257-270.

Rajaraman, A., & Ullman, J. D. (2011). *Mining of massive datasets*, 2012.

Sobehart, J. R., Keenan, S. C., & Stein, R. (2000). Benchmarking quantitative default risk models: a validation methodology. *Moody's Investors Service*.

Stein, R. M. (2002). Benchmarking default prediction models: Pitfalls and remedies in model validation. *Moody's KMV, New York*, 20305.

Van Thiel, D., & van Raaij, F. (2017). Targeting the robo-advice customer: The development of a psychographic segmentation model for financial advice robots. *Journal of Financial transformation, Capco Institute*, 46, 88-104.
Van Thiel, D., & Van Raaij, F. (2017). Explaining customer experience of digital financial advice. *Economics*, 5(1), 69-84.
Walker, E. (1996). Measurement, Regression and Calibration. *Journal of the American Statistical Association*, 91(433), 434-436.
Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). Data Mining: Practical machine learning tools and techniques. *Morgan Kaufmann.*
Wu, X., Zhu, X., Wu, G. Q., & Ding, W. (2014). Data mining with big data. *IEEE Transactions on Knowledge and Data Engineering*, 26(1), 97-107.
Zhang, Y., Jia, H., Diao, Y., Hai, M., & Li, H. (2016). Research on credit scoring by fusing social media information in online peer-to-peer lending. *Procedia Computer Science*, 91, 168-174.
Zion Market Research. (2018). Global digitization in lending. Retrieved from https://www.zionmarketresearch.com/sample/digitization-in-lending-market