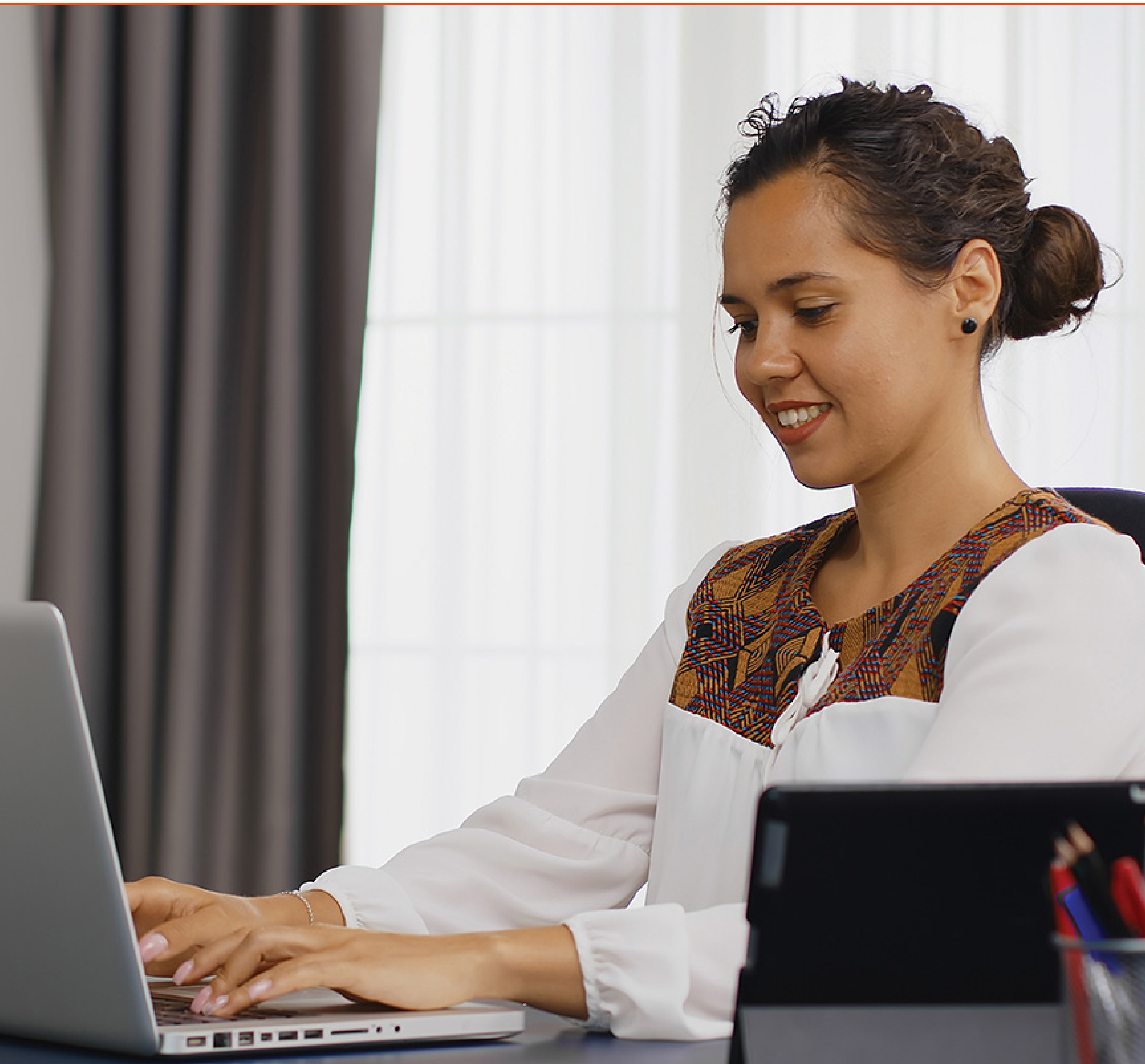
Guide to Mitigating Credit Risk



INTRODUCTION

Risk assessment is crucial for any enterprise that extends credit to customers. Commonly known as credit scoring, the process helps lenders make confident, informed decisions on whether prospective customers will honor their debt. Credit scoring is typically associated with the banking and financial service sectors, but is required across a wide array of businesses, including telecoms, retail, and insurance. In most cases, credit scoring isn't just a business tool, it's a regulatory necessity. And credit scoring is a vast industry. In the U.S. alone, recent consumer debt valuations hover over $14 trillion.

Credit scoring is a complex task that involves wrangling a diverse range and large volume of data. Based on predictive modeling, the use of artificial intelligence (AI) and machine learning (ML) is well established and widespread. And because of the data burden, credit risk firms were some of the earliest organizations to adopt the technology. As such, the credit risk sector can claim to be a pioneer in AI and ML utilization.

Credit Scoring Using Data Analytics

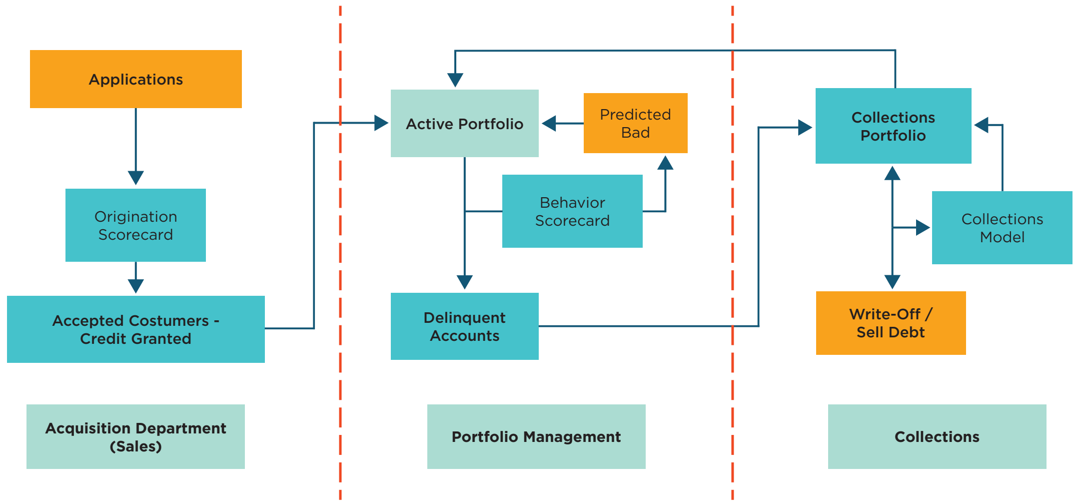
This guide introduces users to the main steps involved in developing credit scoring models. We also review some key techniques and tools available that make this challenging work easier and more efficient. Of course, each enterprise will be driven by its own commercial strategies and priorities. However, there are basic principles that assist any user, regardless of industry or position. By exploiting an individual customer's data, and that of their peers, credit scoring predicts the probability of a certain behavior and/or outcome.

What To Expect

The Credit Lifecycle

The business decisions credit scoring informs are more nuanced than a simple "yes/no" response to the question of whether a firm should extend credit to a customer, since credit scoring can also determine loan size, term length, interest rate, and more.

The use of credit scoring also extends much further than risk assessment of new applications (the Origination or Application Scorecard). The graphic below shows the scope of the full credit lifecycle.



Once a firm grants a loan, responsibility shifts from sales (the left-hand section of the diagram) to account management (the center section). In this part of the lifecycle, a behavior scorecard is generated. This predicts an existing customer's default risk, thereby informing account management decisions over the loan's duration. If a customer defaults on a loan (usually by being more than 90 days overdue on a payment), a collection model will be required (in the right-hand section). This assesses how likely a customer is to pay back their debt, and helps the issuer decide whether to pursue recovery measures or write it off.

THE STANDARD SCORECARD MODEL

Over the years, firms have used several different modeling techniques to support credit scoring. However, the technique widely called Standard Scorecard now dominates. Based on logistic regression as the underlying model (where the probability of a certain outcome is derived from one or more variables) nearly 90% of scorecard developers employ the method. Using this method, scorecards are easy to understand and can be implemented quickly. Crucially, this approach delivers an effective, efficient route to compliance with Basel II regulations.

Intuitive and easy to apply, the basic principles of Standard Scorecard are straightforward. A set of variable attributes (customer characteristics) such as age, income, and home ownership are identified. Weighted scores (points) are then calculated for each of these attributes. The sum of these points provides the credit score. A simple example is shown below.

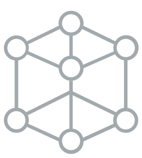
Scorecard Criteria	Range	Points
Age	Up to 25	10
	26 to 40	25
	41 to 65	38
	66 and up	43
Income	Up to 20k	-10
	21k to 40k	16
	41k to 70k	28
	71k and up	45
Bureau Score	Up to 300	-25
	300 - 500	0
	500 - 650	30
	650 - 750	50
	750 +	70
Total Score		**Sum of Points**

SCORECARD MODELING METHODOLOGY


Business Understanding


Data Understanding


Data Preparation


Modeling


Evaluation


Deployment

The job of the data scientist is to design and develop an accurate, verified, and stable credit risk model. What's more, they need to ensure that other data scientists can assess that model and replicate the steps taken.

To do this, users must adopt a framework methodology and model design. Bear in mind that credit scorecard modeling is multidisciplinary – model design needs to be considered not only from a data science perspective, but also in terms of wider business benefits and the development of a viable software platform.

Each of these disciplines will likely require their own methodology. For the model development, Cross Industry Standard Process for Data Mining (CRISP-DM) is recognized as the leading industry methodology. It's comprised of six major phases:

- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment

These phases provide the foundation for users to identify important factors – and their relationships – within a predictive model. It will also drive the formulation of a series of hypotheses regarding credit risk assessments, and the means to test them. Following the CRISP-DM framework, methods will also be established to replicate and validate findings, and gain greater confidence in the model's rigor.

The key building blocks for this journey are:

- The dependent variable (criterion) - for example, the credit status of the applicant
- The candidate independent variables (predictors) - for example, age and income

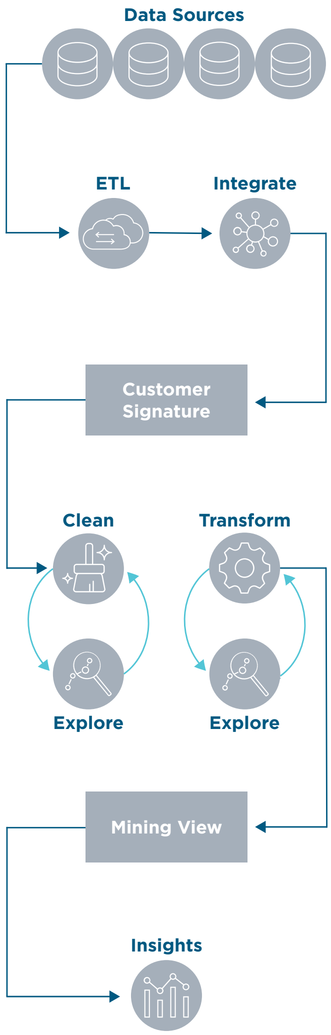
A blueprint for data collection, measurement and analysis is created. This lets users test the model for:

- Reliability – *Are the results stable and consistent?*
- Validity - *Does it represent the phenomenon we're trying to reproduce?*

Good model design should include:

- The unit of analysis – for example, the customer level
- The population – for example, applicants for loans and the sample size
- Operational definitions – for example, what constitutes a "bad" customer?
- The time horizon for observation – for example, a customer's payment history over the last two years
- The performance window – for example, how long does a customer remain "bad?"
- Data sources and data collection methods

DATA PREPARATION AND EXPLORATORY DATA ANALYSIS (EDA)



Data preparation is generally recognized as the most time-consuming and most challenging phase of the CRISP-DM cycle. Typically, users dedicate about 70-90% of total project time to this activity.

Data sources should provide quality. They need to be relevant, accurate, timely, consistent, and complete. But data sources should also offer robust volume and diversity so they provide useful results.

Data Integration, Data Cleansing, and Missing Values

After collection, the next step is data integration and using data merging and concatenation techniques. This is followed by data exploration, otherwise known as exploratory data analysis (EDA). Here, correct interpretation requires an understanding not just of the data itself, but also the wider business context. On this basis, irregularities such as missing values and outliers – which can both jeopardize model accuracy – are dealt with.

Users must exercise considerable caution when dealing with these irregularities. They must ask, *Why is data missing? How can we explain the distribution?* Treatment invariably needs to be tailored to the nature of the problem.

The same is true of outliers. Users must ask, *What is the reason for them? What's their purpose?* It's important to consider that outliers might be helpful for some tasks, like fraud detection.

There's no shortage of statistical and ML tools and techniques to aid the data cleansing process. However, any benefits achieved in terms of greater scorecard accuracy must be valuable enough to justify the extra delay, cost, and complexity these additional data manipulations cause.

Data Transformations

With these tasks complete, we enter a more creative phase. The process of data transformation involves creating additional model variables that are tested for significance. Again, user must combine data science skills with business insight. It's a data scientists' job to suggest the best transformation approach. And very often, the key to creating a good model lies not in the specific modeling technique employed, but the breadth and depth of the independent variables used.

VARIABLE SELECTION

Users should never divorce scorecard modeling from the wider business context. Throughout the journey, “doing more with less” is a core philosophy. To control operating costs, the business should be able to reach loan application decisions as quickly and efficiently as possible. As well as managing risk, enterprises are also under pressure to optimize customer acquisition and improve customer care. Minimizing the number of questions customers need to respond to and then reaching a credit decision quickly are key metrics for enhancing the customer experience. Appropriate variable selection is crucial in achieving these goals.

The previous step, data preparation, created a multi-dimensional customer signature that’s used to discover potentially predictive relationships and then test the strength of those relationships. Business insight analysis is applied at this stage: is the derived customer data in line with business understanding? For example, are customers with a higher debt-to-income ratio more likely to default? Business insight analysis also provides benchmarks for analyzing model results and shapes the modeling methodology.

In terms of selecting variables, the aforementioned “less is more” principle remains paramount. The priority is to identify a minimal set of predictors that will deliver maximum predictive accuracy. Usually, that means a set of somewhere between eight and 15 significant, balanced variables. Unfortunately, variable selection usually isn’t easy. Users might be able to rule out some predictors on legal, ethical, or regulatory grounds, while other variables might remain unidentified.

In practice, selection is a continual process of trial and error. Ultimately, users reach a sweet spot once further significant improvement is impossible. Various measures are available to aid the task; two of the most commonly used measures are information value, for filtering prior to model selection, and stepwise selection, for variable selection (during training of a logistic regression model).

Throughout the journey, “doing more with less” is a core philosophy.

SCORECARD DEVELOPMENT

To turn our data into a scorecard model, the next four steps are:

- Variable transformations
- Model training using logistic regression
- Model validation
- Training

Special variable transformations are essential for any scorecard model based on logistic regression. In layman's terms, this is the process of weighting the variables according to their relative impact on the overall risk of an individual customer defaulting.

The commonly used techniques for variable transformations are:

- Fine classing
- Coarse classing
- Dummy coding or, more usually, weight of evidence (WOE) transformation

Classing is the process by which variables are put into "bins." For example, customer income could be divided into a series of ranges (e.g. $15k-$30k, $30k-$45k, etc.). At the fine classing stage, there might be anything from 20 to 50 of these bins. But again, the "less is more" approach is best. In the coarse classing phase, similar fine grain bins are merged. Typically, just 10 bins emerge.

The overall aim is to simplify the model, minimize information loss, and optimize the variables' predictive power. WOE transformation then assigns each of these coarse classes with a single numeric value. When added together, these values provide the customer's credit score.

Undertaken manually, optimal binning and WOE transformation are time-consuming tasks. Given the need for speed – and the cost of a data scientist's time – automated tools become a stellar business option.

The newly transformed variables need to be checked to ensure they still represent good candidates for the emerging scorecard model. The model must then be scaled to the requirements of the business; the aim here is to provide consistency and score standardization across different scorecards. To simplify interpretation, organizations will often look to use the same score range across multiple scorecards.

Assessment of the model follows, and tests for:

- Accuracy
- Robustness – *How the model will behave when the data changes?*
- Applicability and value – *Is the model valuable from a business perspective?*

SEGMENTATION AND REJECT INFERENCE (RI)

Segmentation

How many scorecards does a business need?

Segmentation should provide the answer. Initially, the process is conducted during the business insights phase, highlighted in Part 3.

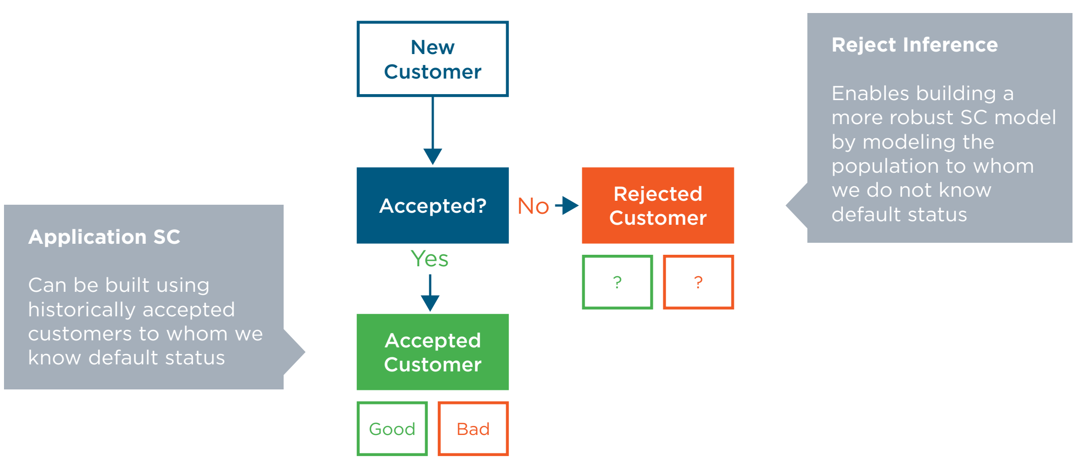
The aim is to identify any population segments with characteristics different enough to warrant their own dedicated scorecard. The need for segmentation might be driven by marketing, with product offerings and target markets that may require new, different models. Data availability might also be an issue.

The “less is more approach” applies here too. To justify their existence, segmented models must offer different predictive patterns and a significant lift in predictive power. In practice, this is rarely the case. To avoid the additional time, cost, and complexity involved, users should make every effort to stick with a single scorecard if they can.

Reject Inference

Reject inference (RI) is a statistical method used to address the naturally occurring bias of modeling that’s based purely on an accepted population with known credit-risk performance. The problem here is that a significant group of customers are excluded from the process because their performance is unknown. To ensure accuracy, scorecard models must include both accepted and rejected populations.

As such, in RI the performance of the “unknowns” must be inferred. Essentially, RI is a form of missing value treatment. Techniques employed include proportional assignment, simple augmentation, fuzzy augmentation, and parcelling.



FURTHER CREDIT RISK MODELING CONSIDERATIONS

Throughout the development journey, the data scientist is aiming for a model that demonstrates all the right characteristics, including rigor, testability, replicability, precision, and confidence. The use of a validation framework, and dealing with unbalanced data, are two more key steps that can help any user achieve this goal.

Overfitting

A phenomenon known as “overfitting” is one of the biggest threats to satisfactory model performance. Essentially, it describes a model that’s too good to be true. The model fits the training data perfectly, but fails to generalize on it. As a result, it offers poor predictive power on new, unseen datasets.

Various validation frameworks exist to both detect and minimize overfitting. Examples include simple validation, nested holdout validation, bootstrapping, and cross-validation.

The Accuracy Paradox

Model accuracy is defined as the ratio of correct predictions to the total number of cases. But this is another area where things aren’t as straightforward as they seem. In part, this is due to the connected issues of unbalanced data and the “accuracy paradox.” The potential impact of the accuracy paradox is illustrated by this simple example: If 1% of the target population is a fraud risk, we can achieve an impressive 99% success rate simply by guessing that each credit application represents “no fraud risk.” Unfortunately, it’ll also mean our miss rate for identifying fraud risk will be a not-so-useful 100%.

Clearly the model needs to strike a balance that considers both correct and incorrect results simultaneously. This means addressing the natural bias towards the majority of users (those who don’t present a fraud risk) and away from the minority of users who users should be worried about (those who do present a fraud risk). Organizations must also consider broader commercial objectives. For example, does the business want to minimize the miss rate for fraud risks or minimize false alarms?

The problem of unbalanced data is addressed by under sampling or oversampling techniques, or a popular modification known as Synthetic Minority Oversampling Technique (SMOTE).

Ultimately, the accuracy paradox means that accuracy isn’t the preferred metric for unbalanced data – the more useful measure is sensitivity, otherwise known as recall, hit rate, probability of detection, true positive rate, true negative rate, or precision.

CREDIT RISK STRATEGIES

Determining a credit risk strategy comes after scorecard development but before implementation. It defines both how the business will interpret the customer's score, and the resulting action organizations will take based on that score. The best credit risk strategies are those that increase the customer base, reduce credit risk, and maximize profit.

The most common strategy is a simple cut-off point for accept/reject, based on a certain customer scorecard value. In addition to a single cut-off, a value range for conditional acceptance can also be included, or a margin in which a member of staff can make a final decision manually. If segmentation is employed, different cut-off values may be set for different customer segments

The role of wider business objectives needs to be stressed. Is the aim to maximize market share by increasing the acceptance rate, or maximize profit by adopting a more risk adverse strategy? Credit risk strategy may also be shaped by customer retention targets, or lifetime customer value calculations. It can also embrace risk-based pricing, reflected in the interest rate, credit limit or repayment term offered to a customer.

Each business will pursue its own path. However, in general terms, there is undoubtedly a danger in adopting an over-simplistic approach that rejects potentially loyal and profitable customers based on a hard yes/no cut-off score.

SCORECARD IMPLEMENTATION: DEPLOYMENT, PRODUCTION AND MONITORING

Implementation is the final stage of the CRISP-DM framework. It represents the transition from the data science domain to the IT domain.

Scorecard implementation is a clearly defined, sequential process:
Create deployment code → Implement and test on a pre-production server → Live launch on a production server → Test scores → Post-production monitoring

Writing deployment code is often error-prone. For many enterprises, it also creates a significant bottleneck. Several code refinement cycles are necessary to produce the finished deployment code. However, some analytics software vendors offer automatic code deployment capability. This produces error-free code and shortens both development time and the code testing cycle.

After testing on the pre-production server, approved models are uploaded to the production environment in either active or passive state:

- Active state: scorecards are used in real-time decision making
- Passive state: scorecards are not typically used in decision making; instead, scores are recorded and analyzed over time to assess their business value.

All models degrade over time. Regular monitoring is therefore essential. Typically, businesses pre-define their performance threshold values, so when performance dips below an acceptable level, remedial action is taken.

When implemented, the model will also give the credit risk department access to a range of reports that provide insight into the factors impacting model performance.

For enterprises, the biggest challenge with model monitoring lies not in knowing they need change, but the time lag between a request for change and its implementation. This is another area where automated model monitoring solutions make a strong business case.

THE BIGGER PICTURE – ENTERPRISE DECISION MANAGEMENT (EDM) SYSTEMS

The development and deployment of a credit scorecard model is challenging, complex, and time-consuming. Yet it represents just one part of a complete enterprise decision management (EDM) system. Other important pieces that make up the bigger picture include:

- Customer application processing
- Internal and external data gathering
- Policy rules
- Additional analytical models for fraud detection, risk management, etc.

Ultimately, the EDM establishes the overarching framework for translating data into actionable decisions; it should bring together ease of implementation, speed of change, and regulatory compliance. As such, it needs to encompass the following capabilities:

- Automation
- Data and system security
- Concurrency of processes
- Scalability
- Transparency
- Heterogeneity with diversity of data sources

Financial institutions and other credit-lending businesses can opt for a completely customized solution. However, this is likely to cost a lot, require complex system maintenance requirements, and place a significant burden on internal staff.

The alternative – a commercial EDM system – promises faster implementation and reduced demand on business resources. As such, commercial EDM systems are quick and cost-efficient. Moreover, organizations can choose solutions that feature intuitive, point-and-click user interfaces so any user can undertake an array of vital processes, including creating decision requirements diagrams, specifying input parameters, controlling model outputs, and implementing business rules – all without needing any technical coding or IT knowledge. In short, this type of approach enables organizations to deploy faster, better scorecard models that address the myriad of demands placed on them in the fiercely competitive modern consumer credit market.

Automation

Data and System Security

Concurrency of Processes

Scalability

Transparency

Heterogeneity with Diversity of Data Sources

Altair is a global leader in computational science and artificial intelligence (AI) that provides software and cloud solutions in simulation, high-performance computing (HPC), data analytics, and AI. Altair enables organizations across all industries to compete more effectively and drive smarter decisions in an increasingly connected world – all while creating a greener, more sustainable future.

For more information, visit www.altair.com