INTELLIGENT CREDIT SCORING

SECOND EDITION

BUILDING AND IMPLEMENTING BETTER CREDIT RISK SCORECARDS

NAEEM SIDDIQI

WILEY

Intelligent Credit Scoring

Building and Implementing Better Credit Risk Scorecards

Second Edition

Naeem Siddiqi

WILEY

Cover image: background © Candice Cusack/iStockphoto; spreadsheet © Pali Rao/ Getty Images, Inc.

Cover design: Wiley

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

Library of Congress Cataloging-in-Publication Data:

ISBN 9781119279150 (Hardcover)
ISBN 9781119282334 (ePub)
ISBN 9781119282297 (ePDF)

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

For Saleha

Acknowledgments

As with the first edition, I am indebted to many people who have provided ideas, advice, and inspiration for the content of this book.

- I would like to thank Dr. Billie Anderson, Dr. Hendrik Wagner, Clark Abrahams, Bradley Bender, and Charles Maner for graciously agreeing to contribute very informative guest chapters to this book.
- The Roman poet Ovid once said, "A horse never runs so fast as when he has other horses to catch up and outpace." I am grateful to the incredibly talented group of people I work with at SAS who continue to enhance my knowledge of risk management and analytics.
- I want to thank Nikolay Filipenkov and Clark Abrahams for reviewing this book and providing excellent ideas for improvements.
- I continue to learn about the contemporary issues in the industry, the challenges, and innovative ways to deal with them from my customers and colleagues in the credit scoring business. What we know today is due in large part to the generous sharing of knowledge and research work done by credit scoring practitioners. We are indebted to them.

My family—Saleha, Zainab, and Noor—who have been incredibly supportive and tolerant of my frequent work-related absences from home, and have done much of the heavy lifting at home during those times (especially during snowstorms in Markham while I am at warmer locales).

Finally, as always, I want to acknowledge my parents for encouraging me to continuously seek knowledge, and for their constant prayers and blessings, without which there would be no success.

Intelligent Credit Scoring

Contents

CHAPTER 1

Introduction

"The only virtue of being an aging risk manager is that you have a large collection of your own mistakes that you know not to repeat."

—Donald Van Deventer

Much has changed since the publication of the first edition of this book in 2006. The use of credit scoring has become truly international, with thousands of lenders now developing their own scorecards in-house. As a benchmark, The SAS Credit Scoring[1] solution, which started out around that time, now has hundreds of customers—but more importantly, they are spread out across 60-plus countries. Many more banks, of course, use products from other vendors to build and use credit risk scorecards in-house, but in general, the trend has moved away from outsourcing the development of scorecards to internal builds. The following factors, listed in the order discussed, have led to more widespread usage of scorecards and the decision by banks to build them in-house.

Factors driving the increased use of scorecards include:

- Increased regulation.
- Ease of access to sizable and reliable data.
- Better software for building scorecards.
- Availability of greater educational material and training for would-be developers.
- Corporate knowledge management fostering retention and sharing of subject-matter expertise.
- Signaling capabilities to external and internal stakeholders.
- Efficiency and process improvement.
- Creating value and boosting profitability.
- Improved customer experience.

In the past decade, the single biggest factor driving banks to bring credit scorecard development in-house has been the Basel II Accord.[2]

Specifically, banks that have opted to (or were told to) comply with the Foundation or Advanced Internal Ratings Based approaches of Basel II were required to internally generate Probability of Default (PD) estimates (as well as estimates for Loss Given Default [LGD] and Exposure at Default [EAD]). Larger banks expanded their production and usage of credit scoring, and were compelled to demonstrate their competence in credit scoring. In many countries, particularly in Europe, even small banks decided to go for these approaches, and thus had to start building models for the first time. This led to some challenges—when you have never built scorecards in-house (and in some cases, not really used them either), where do you start? Many institutions went through significant changes to their data warehousing/management, organizational structure, technology infrastructure, and decision making as well as risk management cultures. The lessons from some of these exercises will be shared in chapters on creating infrastructures for credit scoring, as well as the people who should be involved in a project.

While there is a lot of variance in the way Basel II has been implemented in Europe, it is largely a finished process there.[3] Some of the lessons, from Basel II, specifically on how the default definition should be composed will be detailed in a guest chapter written by Dr. Hendrik Wagner. The implementation of Basel II is still ongoing in many countries, where the same exercise is being repeated many times (and in most cases, the same questions are being asked as were 10 years ago in Europe). Many institutions, such as retail credit card and automotive loan companies, that were not required to comply with Basel II, voluntarily opted to comply anyway. Some saw this as a way to prove their capabilities and sophistication to the market, and as a seal of approval on the robustness of their internal processes. But the ones who gained most were those who saw Basel II compliance not just as a mandatory regulatory exercise, but rather as a set of best practices leading to an opportunity to make their internal processes better. This theme of continuous improvement will be addressed in various parts of the book, and guidance given on best practices for the scorecard development implementation process.

In some countries where Basel II was not a factor, local banks decided to take on analytics to improve and be more competitive. In many developing countries, the banking industry became deregulated or more open, which allowed international banks to start operating there. Such banks generally tended to have a long history of using advanced analytics and credit scoring. This put competitive pressures on some of the local banks, which in many cases were operating using manual and judgmental methods. The local banks thus started investing in initiatives such as data warehousing, analytics, and in-house credit scoring in order to bring costs down, reduce losses, and create efficiencies. Another factor that points to a wider acceptance of credit scoring is the tight market for scorecard developers globally. In almost all the countries, whether those with Basel II or not, the demand for experienced credit scoring resources has continued to be high.

In more recent times, the introduction of International Financial Reporting Standards (IFRS) 9 to calculate expected losses has expanded the usage of predictive models within all companies. Those institutions that have already invested in fixing their data problems and establishing sustainable and robust analytics functions will find it easier to comply.

In mature markets, banks that had been developing models and scorecards before have now been looking at how to make the process efficient, sustainable and more transparent. Investments in data warehousing, tools to enable analysts to access the data quickly and easily, integrated infrastructure to reduce model risk, governance processes, and other such areas have increased. Many banks that had invested a lot of money into data warehousing were also looking to increase return on investment (ROI). Credit scoring offered a quick and proven way to use the data, not just for reducing losses but also lead to greater profitability.[4]

Scarcity of modeling/credit scorecard (these two words are used interchangeably throughout this book) development resources has led institutions to try to reduce human resources risk by using modeling tools that encourage sharing and retention of corporate knowledge, reduce training cycles and costs, and are easier to use. Some of the challenges and risks of developing scorecards in-house will be discussed in the chapter on managing the risks of in-house scoring.

In other banks not specifically impacted by the preceding, increasing competition and growing pressures for revenue generation have led credit-granting institutions to search for more effective ways to attract new creditworthy customers and, at the same time, control losses. Aggressive marketing efforts have resulted in a continuously deeper penetration of the risk pool of potential customers, and the need to process them rapidly and effectively has led to growing automation of the credit and insurance application and adjudication processes. The risk manager is challenged to produce risk adjudication solutions that can not only satisfactorily assess creditworthiness but also keep the per-unit processing cost low, while reducing turnaround times for customers. In some jurisdictions without a credit bureau, the risk manager faces an additional challenge of doing so using data that may not be robust or reliable. In addition, customer service excellence demands that this automated process be able to minimize denial of credit to creditworthy customers, while keeping out as many potentially delinquent ones as possible.

At the customer management level, companies are striving ever harder to keep their existing clients by offering them additional products and enhanced services. Risk managers are called on to help in selecting the "right" (i.e., low-risk) customers for these favored treatments. Conversely, for customers who exhibit negative behavior (nonpayment, fraud), risk managers need to devise strategies to not only identify them but also to deal with them effectively to minimize further loss and recoup any monies owed as quickly as possible.

It is in this environment that credit risk scorecards have continued to offer a powerful, empirically derived solution to business needs. Credit risk scorecards have been widely used by a variety of industries for predicting various types of payment delinquencies, fraud, claims (for insurance), and recovery of amounts owed for accounts in collections, among other things. More recently, as mentioned previously, credit scoring has been used widely for regulatory compliance. Credit scoring offers an objective way to assess risk, and also a consistent approach, provided that system overrides are maintained below acceptable policy-specified thresholds.

In the past, most financial institutions acquired credit risk scorecards from a handful of credit risk vendors. This involved the financial institution providing their data to the vendors, and the vendors then developing a predictive scorecard for delivery. For smaller companies, buying a generic or pooled data scorecard was the only option. While some advanced companies have had internal modeling and scorecard development functions for a long time, the trend toward developing scorecards in-house has become far more widespread in the past few years. Some of the regulatory and operational reasons for this phenomenon were covered at the beginning of this chapter. Others will be discussed later.

First, there are more powerful and easy-to-use data mining software today than ever before. This has allowed users to develop scorecards without investing heavily in advanced programmers and infrastructure. Growing competition and the entry of several new data mining vendors made such tools available at ever cheaper prices. Complex data mining functions became available at the click of a mouse, allowing the user to spend more time applying business and data mining expertise to the problem, rather than debugging complicated and lengthy programs. The availability of powerful "point-and-click"–based Extract-Transform-Load (ETL) software enabled efficient extraction and preparation of data for scorecard development and other data mining. Second, advances in intelligent and easy-to-access data storage have removed much of the burden of gathering the required data and putting it into a form that is amenable to analysis. As mentioned earlier, banks and other lenders have made significant investments in data warehousing and data management, and are now looking to use that data to increase profitability.

Once these tools became available, in-house development became a viable option for many smaller and medium-sized institutions. The industry could now realize the significant ROI that in-house scorecard development could deliver for the right players. Experience has shown that in-house credit scorecard development can be done faster, cheaper, and with far more flexibility than any outsourcing strategy. Development was cheaper since the cost of maintaining an in-house credit scoring capability was less than the cost of purchased scorecards. Internal development capability also allowed companies to develop far more scorecards (with enhanced segmentation) for the same expenditure. Scorecards could also be developed more rapidly by internal resources using the right software—which meant that better custom scorecards could be implemented more rapidly, leading to lower losses.

In addition, companies have increasingly realized that their superior knowledge of internal data and business insights led them to develop better-performing scorecards. Seasoned modelers understand that the single biggest contributor to model quality is the data itself, followed by the knowledge level of the analyst of that data. This book will cover in detail how internal knowledge can be applied to build better scorecards. In every phase of the project, we will discuss how appropriate judgment can be applied to augment statistical analyses.

Better-performing scorecards also came about from having the flexibility to experiment with segmentation and then following through by developing more finely segmented scorecards. Deeper segmentation allows for more fine-tuned predictions and strategies. Combined with software that can implement champion/challenger scorecards, this becomes a great way to experiment with different configurations of models. Performing such detailed segmentation analysis through external vendors can become expensive.

Banks have also realized that credit risk scorecards are not a commodity to be purchased from the lowest bidder—they are a core competence and knowledge product of the institution. Internal scorecard development increases the knowledge base within organizations. The analyses done reveal hidden treasures of information that allow for better understanding of customers' risk behavior and lead to better strategy development. We will cover some of this knowledge discovery in the section on model development, specifically the grouping process.

In summary, leaving key modeling and strategy decisions to "external experts" can prove to be a suboptimal route at best, and can also be quite costly.

This book presents a business-focused process for the development and usage of credit risk prediction scorecards, one that builds on a solid foundation of statistics and data mining principles. Statistical and data mining techniques and methodologies have been discussed in detail in various publications and will not be covered in depth here. I have assumed that the reader is either familiar with these algorithms, or can read up on them beforehand, and is now looking for business knowledge pertaining to scorecard development.

The key concepts that will be covered in the book are:

- The application of business intelligence to the scorecard development process, so that the development and implementation of scorecards is seen as an intelligent business solution to a business problem. Good scorecards are not built by passing data solely through a series of programs or algorithms—they are built when the data is passed through the analytical and business-trained mind of the user. This concept will be applied in all the sections of this book—taking statistical analyses and overlaying business knowledge on it to create better results.

- Building scorecards is a business process—as much as we use statistical algorithms, simple or complex, to build models, at the end of the day it is a business exercise. The purpose of the exercise is to enable a better business decision and not merely the creation of a great formula. As such, each process—whether selecting a "bad" definition, deciding appropriate segmentations, best bins for attributes, or the best scorecard—will be viewed through the lens of a business decision.
- Collaborative scorecard development, in which end users, subject matter experts, implementers, modelers, validators, decision makers and other stakeholders work in a cohesive and coherent manner to get better results and avoid costly setbacks and potential disasters during the process.
- The concept of building a risk profile—this means building scorecards that contain predictive variables representing major information categories, usually between 8 and 15 variables. This mimics the thought processes of good risk adjudicators, who analyze information from credit applications or customer behavior and create a profile based on the different types of information available. They would not make a decision using four or five pieces of information only—so why should anyone build a scorecard that is narrow based? In statistics, parsimonious models are usually preferred. However, in this case, where the modeler is attempting to more fully capture the business reality, more variables are preferred in order to construct a proper and representative risk profile. The point of the exercise is to make the best decision-making tool possible, not just a statistical one.
- Anticipating impacts of decisions and preparing for them. Each decision made—whether on the definition of the target variable, segmentation, choice of variables, transformations, choice of cutoffs, or other strategies—starts a chain of events that impacts other areas of the company as well as future performance. By tapping into corporate intelligence and working in collaboration