Credit Scoring in the Age of Big Data

Billie Anderson
Bryant University, USA

J. Michael Hardin
University of Alabama, USA

INTRODUCTION

Credit scoring is a method of modeling potential risk of credit applicants. It involves using different statistical techniques and past historical data to create a credit score that financial institutions use to assess credit applicants in terms of risk. Credit scoring is essentially a type of classification problem: which credit applicants should be considered good risks and which applicants should be considered bad risks.

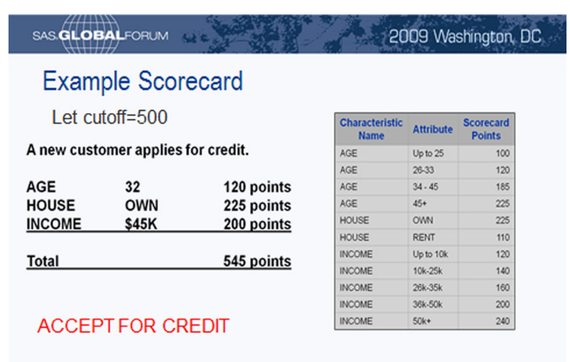
A scorecard model is built from a number of characteristic inputs. Each characteristic is comprised of a number of attributes. In the example scorecard shown in Figure 1, age is a characteristic and "25–33" is an attribute. Each attribute is associated with a number of scorecard points. These scorecard points are statistically assigned to differentiate risk, based on the predictive power of the variables, correlation between the variables, and business considerations.

For example, in Figure 1, the credit application of a 32 year old person, who owns his own home and makes $30,000, would be accepted for credit by this institution. The total score of an applicant is the sum of the scores for each attribute present in the scorecard. Lower scores imply a higher risk of default, and higher scores imply lower risk.

Within the past few years, the development of an accurate credit scoring model has become a priority for several reasons: growth in competition among credit card companies, a rising number of bad loans as a result of a weak United States economy, and the need for more stringent government regulations. The recent Congressional Dodd-Frank Act added an additional 250 regulations that will involve 11 governmental bodies. These challenges have prompted the financial industry to explore and use non-traditional data sources as part of the loan granting decision.

For years the financial industry has managed large volumes of data generated by customer, operational, and regulatory sources. Thus, banks are thoroughly familiar with big data – massive amounts of unstructured data. Financial service lenders are the most data intensive economic sector (Rubin, 2011). There is a shift taking place as to how individuals interact with their bank. Many customers are turning to digital channels to conduct transactions rather than use the traditional face-to-face branch relationship. Banks must now respond and use the data collected from digital sources to make real-time loan offers with the highest acceptance rate possible. Banks are learning how to use big data sources to monitor changes in customer behavior and to improve the banking experience instantaneously. More data is now available than ever before; the challenge for

Figure 1. Example scorecard



DOI: 10.4018/978-1-4666-5202-6.ch049

financial institutions will be how to put that data to work and make the smartest lending decisions.

This chapter will describe how banks and financial organizations are starting to incorporate big data sources, such as data from social media websites, into the credit lending process. A discussion of how more established organizations, such as Experian and SAS, are incorporating big data in their scorecard methodology will be given. A description of two start-up companies which are using electronic and big data sources such as social media exclusively to provide banking services and grant loans will be discussed.

BACKGROUND

The statistical methods used to categorize objects into groups can be traced to 1936 in Fisher's publication (Fisher, 1936). Durand (1941) was the first to use Fisher's methodology to distinguish between good and bad loans. Using this research, the founders of credit scoring, Bill Fair and Earl Isaac, built the first credit scoring system for the United States in 1958. Although credit scoring has been in use since that time, it is only recently that credit scoring has become widespread.

Abdou & Pointon (2011) provide the most contemporary and comprehensive review of the credit scoring literature. The authors surveyed 214 articles, books and manuscripts related to credit scoring applications in the business field. The authors conclude that different types of credit scoring models are applicable under different situations. There are some older, yet still very relevant credit scoring resources that discuss the statistical issues in developing a credit scorecard (Siddiqi, 2006; Thomas, Oliver, & Hand, 2005; Hand & Henley, 1997).

As with any other business sector that is highly data dependent, the credit scoring world has felt the impact of the big data phenomenon that is sweeping through modern businesses (Lohr, 2012). Banks are moving away from using traditional statistical techniques to build a credit scorecard. Financial organizations are combining the traditional techniques with new big data technology and big data analytics to build their scorecards. With the rise of companies using social media data as a tool to build better predictive models, financial institutions are integrating big data sources in calculating credit risk (Liew, 2012). Recently a start-up company has raised $8 million to be distributed in the forms of loans for which the applicants are approved based on their online social reputation (Gage, 2012).

MAIN FOCUS

In the last year, financial institutions have begun exploring ways to incorporate an applicant's social media activity such as Facebook and Twitter into the risk associated with a loan. The main advantage for a financial institution using such expanded data sources is the shorter time frame to deploy a credit scoring model. In a traditional credit scoring scenario, it can take 12 to 18 months to update a scorecard model with new data. This time lag results from lenders having to verify the new data, determining if the new data impacts other departments of the institution, and taking time for employees to learn how to use the new information to make decisions on credit worthiness, pricing and cross-selling (Adams, 2012).

By using big data sources and the associated big data analytics, a bank can increase:

- **Efficiency:** Efficiency is essential to banks that collect and analyze vast volumes of data. Accessing and analyzing various types of structured and unstructured data as soon as it is captured is essential in staying relevant to electronic banking customers.
- **Profitability:** Big data sources allow a bank to reach customers in ways that were never possible. By using online banking and social media applications, a bank can