

Statistica Sinica Preprint No: SS-2017-0043.R1

Title	RFMS Method for Credit Scoring Based on Bank Card Transaction Data
Manuscript ID	SS-2017-0043.R1
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202017.0043
Complete List of Authors	Jing Zhou Danyang Huang and Hansheng Wang
Corresponding Author	Jing Zhou
E-mail	zhoujing_89@126.com
Notice: Accepted version subject to English editing.	

the approval of applicants, their transaction history can be traced using the registered phone number. Compared to traditional data acquisition method (e.g., customer self-reported data), this data source has some unique advantages by avoiding three problems as follows: (1) Over-simplified portrayal of individual who lacks of a multidimensional evaluation; (2) High risk of financial fraud such as exaggeration of employee payment which is impossible for companies to detect; (3) Lack of cross validation that impacts the reliability of credit evaluation.

If an applicant is rejected by the company and thus no loan is approved, he should then have no chance to observe its default behavior (i.e., the response). As a result, those applicants cannot be used for modeling. Accordingly, we should focus on those applicants, whose loans have been actually approved. However, by the time of maturity, some of them successfully return the principal and interest, while the others fail. For those applicants, we have their complete transaction records and thus can be used for modeling. This might leads to the bias of credit evaluation and is a challenging problem faced by the entire industry. In our work, we follow the industry common practice and focus on approved applicants only. This leads to a sample of 26,513 applicants. There are almost 1500,000 transaction records and other related information over a year. The data contains 7,980 default applicants and 18,533 non-default ones. It is worth noting that this is a sample for model analysis and thus it does not reflect the real default rate among the applicants who adopt the product.

For each applicant, we can collect four types of information, they are applicant information, merchant classification, bank card information and transaction records. See Figure 1 for illustration. These four sources of information can be linked by different keywords as follows: (1) Applicant information and transaction records are connected by phone number. Each phone number corresponds to several transaction records. (2)

