

Nghiên cứu một số kỹ thuật khai phá dữ liệu cho bài toán dự đoán rủi ro tín dụng :

Luận văn ThS / Nguyễn Văn Toàn ; Nghd. : TS.

Nguyễn Đức Dũng . - H. : ĐHCN, 2008 . - 82 tr. + CD-ROM

Tóm tắt: Tổng quan khai phá tri thức từ dữ liệu và khai phá dữ liệu; trình bày bài toán dự đoán rủi ro tín dụng, thông tin về khoản vay và dữ liệu sử dụng cho bài toán; trình bày xử lý dữ liệu dự đoán rủi ro tín dụng và dự đoán rủi ro tín dụng bằng sử dụng cây quyết định và sử dụng phương pháp quy nạp luật; giới thiệu Support Vector Machine, đặc điểm và dự đoán rủi ro tín dụng bằng Support Vector Machine, đưa ra những đánh giá kết quả dự báo rủi ro tín dụng bằng SVM

MỞ ĐẦU	5
CHƯƠNG 1 - TỔNG QUAN KHAI PHÁ TRI THỨC TỪ DỮ LIỆU VÀ KHAI PHÁ DỮ LIỆU	12
1.1 Giới thiệu khai phá tri thức từ dữ liệu và khai phá dữ liệu	12
1.2 Nhiệm vụ(task) khai phá dữ liệu	15
1.3 Đánh giá mô hình dự đoán	16
1.4 ứng dụng KDD và Datamining	18
CHƯƠNG 2 - BÀI TOÁN DỰ ĐOÁN RỦI RO TÍN DỤNG 21	21
2.1. Bài toán dự đoán rủi ro tín dụng.....	21
2.2. Thông tin về một khoản vay.....	22
2.2.1. Thông tin về khách hàng vay vốn.....	22
2.2.2. Thông tin về phương án sản xuất kinh doanh	23
2.2.3. Thông tin liên quan đến tài sản bảo đảm (TSBD)	24
2.2.4. Thông tin khác.....	24
2.3. Dữ liệu sử dụng cho bài toán.....	24
CHƯƠNG 3 - XỬ LÝ DỮ LIỆU DỰ ĐOÁN RỦI RO TÍN DỤNG 25	25
3.1. Tiền xử lý dữ liệu	25
3.1.1. Công việc tiền xử lý dữ liệu	25
3.1.2. Tiền xử lý dữ liệu dự đoán rủi ro tín dụng	30
3.2. Phương pháp lấy mẫu dữ liệu (Data sampling)	35
3.2.1. Phương pháp lấy mẫu dữ liệu.....	35
3.2.2. Lấy mẫu lên (Oversampling)	35
3.2.3. Lấy mẫu xuống (undersampling)	36
3.2.4. Kết hợp hai chiến lược trên.....	36
3.3. Hợp nhất các lớp của tập dữ liệu (Data merged).....	36
CHƯƠNG 4 - DỰ ĐOÁN RỦI RO TÍN DỤNG SỬ DỤNG 38	38
CÂY QUYẾT ĐỊNH (C4.5) 38	38
4.1. Giới thiệu cây quyết định	38
4.1.1. Khái niệm.....	38
4.1.2. Các loại cây quyết định	38
4.1.3. Ví dụ thực hành	39

4.1.4. Xây dựng cây quyết định	40
4.2. Thuật toán C4.5 và đặc điểm cây quyết định	41
4.2.1. Thuật toán C4.5 [17]	41
4.2.2. So sánh C4.5 với các phiên bản khác	45
4.2.3. Đặc điểm cây quyết định	46
4.3. Dự đoán rủi ro tín dụng sử dụng cây quyết định (C4.5).....	46
4.4. Đánh giá kết quả dự đoán rủi ro tín dụng sử dụng cây quyết định (C4.5).....	52
CHƯƠNG 5 - DỰ ĐOÁN RỦI RO TÍN DỤNG SỬ DỤNG PHƯƠNG PHÁP QUY	

NẠP LUẬT 58

5.1. Giới thiệu.....	58
5.2. Thuật toán CBA	59
5.3. Dự đoán rủi ro tín dụng bằng CBA.....	65
5.4. Đánh giá kết quả dự đoán rủi ro tín dụng.....	67

CHƯƠNG 6 - Dự ĐOÁN RỦI RO TÍN DỤNG **sử** DỤNG SUPPORT VECTOR

MACHINE 71

6.1. Giới thiệu Support Vector Machine.....	71
6.2. Đặc điểm của Support Vector Machine	73
6.3. Dự đoán rủi ro tín dụng bằng Support Vector Machine.....	75
6.3.1. Lựa chọn tham số.....	75
6.3.2. Thực nghiệm với tập dữ liệu DI	76
6.3.3. Thực nghiệm với tập dữ liệu được hợp nhất lớp (D2)	77
6.4. Đánh giá kết quả dự đoán rủi ro tín dụng bằng SVM	78

CHƯƠNG 7 - ĐÁNH GIÁ CÁC KỸ THUẬT ĐỐI VỚI BÀI TOÁN Dự ĐOÁN RỦI

RO TÍN DỤNG.....	79
TÀI LIỆU THAM KHẢO	82