



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2009-0006489
(43) 공개일자 2009년01월15일

(51) Int. Cl.

G06Q 10/00 (2008.03)

(21) 출원번호 10-2007-0069863

(22) 출원일자 2007년07월11일

심사청구일자 2007년07월11일

(71) 출원인

고려대학교 산학협력단

서울 성북구 안암동5가1 고려대학교 내

(72) 발명자

박유성

서울 성북구 안암동5가 1번지 고려대학교

이동희

충남 연기군 조치원읍 고려대학교 서창캠퍼스

(뒷면에 계속)

(74) 대리인

현중철

전체 청구항 수 : 총 14 항

(54) 신용 위험 모형 구축용 툴 킷, 신용 위험 모형 구축을 통한 신용 위험 관리 방법 및 이를 기록한 기록매체

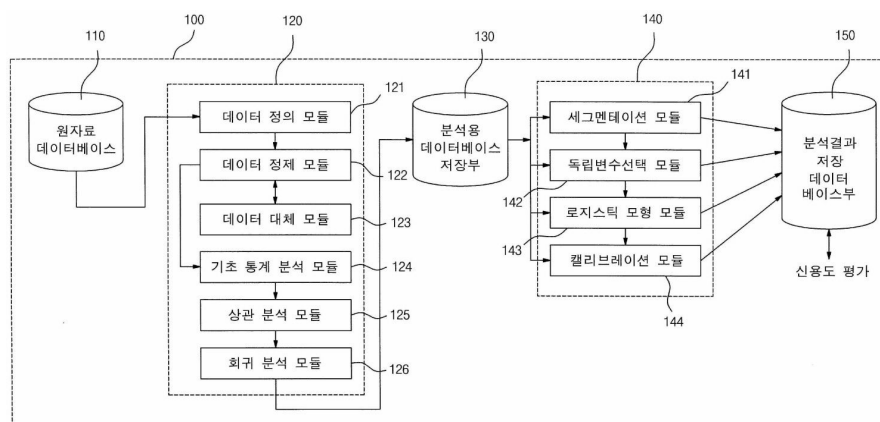
(57) 요약

신용 위험 모형 구축용 툴 킷, 신용 위험 모형 구축을 통한 신용 위험 관리 방법 및 이를 기록한 기록매체가 개시된다.

본 발명에 따른 신용 위험 모형 구축용 툴 킷은 신용도 평가를 위하여 기업 또는 개인의 재무제표에 대한 데이터를 저장하는 원자료 데이터베이스; 상기 원자료 데이터베이스로부터 데이터를 선택하여 상기 선택된 데이터의 결측치를 제거하고, 상기 결측치가 제거된 데이터를 관리하는 자료 관리부; 상기 자료 관리부의 데이터를 저장하는 분석용 데이터베이스 저장부; 상기 분석용 데이터베이스 저장부의 데이터로부터 신용도 평가 모형의 변수를 선택하고, 상기 선택된 변수로부터 로지스틱 회귀분석에 의해 데이터로부터 신용도 평가 모형을 구축하는 모형 구축부; 상기 모형 구축부로부터 구축된 신용도 평가 모형의 분석 결과를 저장하는 분석 결과 저장 데이터베이스부; 및 상기 분석 결과 저장 데이터베이스부를 이용하여 상기 기업 또는 개인의 신용도를 평가하는 모형 평가부를 포함한다.

본 발명에 의하면, 신용 위험관리의 분석을 위하여 현재 보유하고 있는 데이터를 특수한 형태로 변형할 필요가 없으며, 사용자의 상황에 따라 다양한 유형의 문서 데이터를 이용할 수 있고, 작은 규모의 신용 위험관리를 위한 툴 킷으로 대용량의 데이터를 신속하게 처리할 수 있으며, 이상치 및 결측치 등과 같은 복잡한 작업을 그래픽 유저 인터페이스 환경 하에서 간편하게 처리할 수 있고, 변수 선택 분석방법에 따라 신용 위험관리를 위한 다양한 모형을 제공하며, 현재 많이 사용되는 부도예측 모형에 따른 호환성을 제공하고, 기업의 부도 예측을 위한 효과적인 기업의 분류규칙을 제공할 수 있는 효과가 있다.

대표도



(72) 발명자

최보승

서울 성북구 안암동5가 1번지 고려대학교

김태영

서울 동작구 흑석동 186-20번지 삼익빌라 B동 2호

특허청구의 범위

청구항 1

신용도 평가를 위하여 기업 또는 개인의 재무제표에 대한 데이터를 저장하는 원자료 데이터베이스;
 상기 원자료 데이터베이스로부터 데이터를 선택하여 상기 선택된 데이터의 결측치를 제거하고, 상기 결측치가 제거된 데이터를 관리하는 자료 관리부;
 상기 자료 관리부의 데이터를 저장하는 분석용 데이터베이스 저장부;
 상기 분석용 데이터베이스 저장부의 데이터로부터 신용도 평가 모형의 변수를 선택하고, 상기 선택된 변수로부터 로지스틱 회귀분석에 의해 데이터로부터 신용도 평가 모형을 구축하는 모형 구축부;
 상기 모형 구축부로부터 구축된 신용도 평가 모형의 분석 결과를 저장하는 분석 결과 저장 데이터베이스부; 및
 상기 분석 결과 저장 데이터베이스부를 이용하여 상기 기업 또는 개인의 신용도를 평가하는 모형 평가부를 포함하는 신용 위험 모형 구축용 툴 킷.

청구항 2

제 1 항에 있어서,
 상기 자료 관리부는,
 상기 선택된 데이터에서 상기 신용도 평가에 영향을 주는 변수인 독립변수와 상기 독립변수에 따른 신용도 평가의 결과를 나타내는 변수인 반응변수를 지정하고, 상기 독립변수 및 반응변수를 이용하여 상기 선택된 데이터의 기초통계량 정보를 산출하는 데이터 정의 모듈;
 상기 선택된 데이터를 상기 독립변수 내에서 결측치를 검출하여 데이터를 정제하는 데이터 정제 모듈;
 상기 데이터 정제 모듈에서 검출된 상기 결측치를 대체 데이터로 대체하는 데이터 대체 모듈;
 상기 데이터 정제 모듈에서 정제된 데이터를 기반으로 기초 통계량을 분석하는 기초 통계 분석 모듈;
 상기 기초 통계 분석 모듈로부터 분석된 기초 통계량을 기반으로 상기 독립변수 간의 상관 관계를 분석하여 상기 신용도 평가와 관련된 상관 독립변수를 추출하는 상관 분석 모듈; 및
 상기 상관 분석 모듈로부터 추출된 상관 독립변수를 이용하여 상기 반응변수와의 통계적 유의성을 검증하는 회귀분석 모듈을 포함하는 것을 특징으로 하는 신용 위험 모형 구축용 툴 킷.

청구항 3

제 2 항에 있어서,
 상기 모형 구축부는,
 상기 분석용 데이터베이스 저장부에 저장된 데이터를 이용하여 미리 정의된 그룹별로 상기 기업 또는 개인을 분류하는 세그멘테이션 모듈, 상기 분석용 데이터베이스 저장부에 저장된 데이터를 이용하여 상기 신용도 평가를 위해 필요한 독립 변수를 선택하는 독립 변수 선택 모듈, 및 상기 독립 변수를 이용하여 신용도 평가를 위한 컷 오프 포인트를 추출하여 신용도를 예측하는 로지스틱 모형 모듈을 포함하는 것을 특징으로 하는 신용 위험 모형 구축용 툴 킷.

청구항 4

제 3 항에 있어서,
 상기 모형 구축부는 상기 로지스틱 모형 모듈에서 예측된 신용도를 비모수 회귀모형을 이용하여 역회귀분석을 수행하는 캘리브레이션 모듈을 더 포함하는 것을 특징으로 하는 신용 위험 모형 구축용 툴 킷.

청구항 5

제 1 항에 있어서,

상기 원자료 데이터베이스, 분석용 데이터베이스부 및 분석 결과 저장 데이터베이스부에 저장된 데이터는 XML 문서 형식으로 저장되는 것을 특징으로 하는 신용 위험 모형 구축용 툴 킷.

청구항 6

기업 또는 개인의 신용도 평가를 위해 산출된 데이터가 저장된 원자료 데이터베이스로부터 분석하고자 하는 데이터를 선택하는 단계;

상기 선택된 데이터에서 상기 신용도 평가에 영향을 주는 변수인 독립변수와 상기 독립변수에 따른 신용도 평가의 결과를 나타내는 변수인 반응변수를 지정하고, 상기 독립변수 및 반응변수를 이용하여 상기 선택된 데이터의 기초통계량 정보를 산출하는 단계;

상기 선택된 데이터 중 상기 독립변수 내에서 결측치를 검출하여 데이터를 정제하는 단계;

상기 정제된 데이터를 기반으로 기초통계량, 상기 독립변수 간의 상관도 및 회귀분석을 수행하여 상기 신용도 평가를 수행할 독립변수를 선별하는 단계;

상기 선별된 독립변수를 이용하여 신용도 평가에 관한 예측 테이블을 구성하고, 로지스틱 회귀분석에 따라 상기 구성된 예측 테이블을 이용하여 신용도 평가 모형을 구축하는 단계; 및

상기 구축된 신용도 평가 모형을 이용하여 상기 신용도 평가값을 산출하여 신용도를 추정하는 단계를 포함하는 신용 위험 모형 구축을 통한 신용 위험 관리 방법.

청구항 7

제 6 항에 있어서,

상기 데이터를 정제하는 단계는,

상기 선택된 독립변수 내에서 결측치가 존재하는 지를 판별하여 결측치가 존재하는 경우, 상기 결측치를 대체 데이터로 대체하는 단계를 포함하고,

상기 대체 데이터는 상기 결측치가 포함되는 독립변수 내에서의 중간값 또는 평균값인 것을 특징으로 하는 신용 위험 모형 구축을 통한 신용 위험 관리 방법.

청구항 8

제 6 항에 있어서,

상기 신용도 평가 모형은 로지스틱 모형을 포함하는 것을 특징으로 하는 신용 위험 모형 구축을 통한 신용 위험 관리 방법.

청구항 9

제 6 항에 있어서,

상기 신용도를 추정하는 단계는,

비모수 회귀모형을 이용하여 상기 신용도 평가값을 재추정하고, 상기 재추정된 신용도 평가값을 이용하여 신용도를 평가하는 단계를 더 포함하는 것을 특징으로 하는 신용 위험 모형 구축을 통한 신용 위험 관리 방법.

청구항 10

제 6 항에 있어서,

상기 구성된 예측 테이블을 이용하여 신용도 평가 모형을 구축하는 단계는,

상기 구성된 예측 테이블을 이용하여 컷 오프 포인트를 추출해내고, 상기 추출된 컷 오프 포인트에 따라 신용도 평가 모형을 구축하는 것을 특징으로 하는 신용 위험 모형 구축을 통한 신용 위험 관리 방법.

청구항 11

제 6 항에 있어서,

상기 기초통계량은,

데이터의 타입, 각 변수의 역할, 데이터 표본의 크기, 결측치의 개수, 평균, 분산 및 상기 데이터의 다섯숫자요약 값을 포함하는 것을 특징으로 하는 신용 위험 모형 구축을 통한 신용 위험 관리 방법.

청구항 12

제 6 항에 있어서,

상기 신용도 평가 모형은

상기 기업 또는 개인의 부도 모형, 특정 금융상품의 고객 구매 모형, 카드회사의 카드 연체 모형 또는 보험가입자의 이탈방지 모형을 포함하는 것을 특징으로 하는 신용 위험 모형 구축을 통한 신용 위험 관리 방법.

청구항 13

제 6 항에 있어서,

상기 구성된 예측 테이블을 이용하여 신용도 평가 모형을 구축하는 단계는,

상기 선별된 독립변수를 이용하여 역회귀분석을 이용하여 신용도 평가 모형을 구축하는 단계를 더 포함하는 것을 특징으로 하는 신용 위험 모형 구축을 통한 신용 위험 관리 방법.

청구항 14

제 6 항 내지 제 13 항 중 어느 한 항의 방법을 실행할 수 있도록 컴퓨터에 저장된 기록매체.

명 세 서

발명의 상세한 설명

기술 분야

- <1> 본 발명은 신용 위험 관리에 관한 것으로서, 더욱 상세하게는 기업 또는 개인의 신용 위험 관리를 위한 모형을 구축하고, 구축된 모형을 기반으로 하여 신용 위험 관리 업무뿐만 아니라 유통분야에 적용할 수 있는 신용 위험 모형 구축용 툴 킷, 신용 위험 모형 구축을 통한 신용 위험 관리 방법 및 이를 기록한 기록매체에 관한 것이다.

배경 기술

- <2> 금융 산업의 발전에 발맞추어 리스크 관리(Risk Management)에 대한 인식이 확산됨에 따라 국제결제기구에 의하여 규제 자본에 대한 필요성이 은행권을 중심으로 대두 되었다. 금융기관들은 리스크 관리의 중요성을 여러 차례의 금융위기를 통하여 인식하였으며, 그 적용기법에 있어서도 많은 발전을 이루고 있다.
- <3> 또한, 산업 기술 분야에도 시장에서의 회사 간 경쟁이 치열해짐에 따라서 기업의 미래 신용도에 대한 예측이 힘들어졌다. 금융기관의 입장에서는 이러한 신용 위험을 얼마나 효율적으로 관리하고 이전할 수 있는가가 경쟁력의 중요한 요인이 되고 있으며, 또한 이에 대한 방법으로 신용 위험에 대한 평가의 필요성이 대두되고 있다.
- <4> 신용 위험이란 자금의 차입자가 약속한 시일에 약속한 금액을 상환하지 못할 가능성을 의미하고, 신용 위험은 신용 등급(Credit Rating), 무위험 이자율(Risk-free Interest Rate)과 같은 기준지표에 대한 스프레드(Spread) 및 부채의 시장가격이 반영되어 나타난다.
- <5> 이와 같은 신용 위험의 관리에 있어서, 국내에서는 소수의 전문화된 기업평가 기관을 통하여 각종 기업의 평가가 수행되고 있으며 주요 금융기관에서는 별도의 신용 위험관리를 위한 시스템의 구축은 시간적, 비용적 측면에서 접근이 어려운 상태이다.
- <6> 그러나, 현재 각 금융기관에서 구축된 신용 위험 관리 모형들은 금융기관의 자체인력에 의한 구현이기보다는 대부분 전문 인력을 통한 외주 위탁(outsourcing)을 통하여 구현되고 있다. 전문 인력들은 전문적인 금융 공학적 지식뿐만 아니라 통계적인 지식을 함께 겸비하고 있어야 하며, 모형 구축을 위한 특화된 시스템보다는 범용적인

통계분석을 위한 프로그램과 상용 개발 소프트웨어를 이용하여 자체적인 개발을 수행하고 있다.

- <7> 따라서, 전문성을 갖춘 신용 위험관리를 위한 시스템을 개발하기 위해서는 상당한 시간과 전문가 활용에 따른 비용이 증가할 수밖에 없는 실정이다.
- <8> 시스템이 구축된 이후에도 시스템 자체의 규모가 상당히 큰 경우가 대부분이다. 따라서 별도의 신용관리 시스템을 운용하기 위해서는 추가적인 하드웨어의 구축 및 운영 시스템의 구축이 수반되어야 하고, 시스템이 구축된 이후에도 적극적인 시스템의 유지 및 활용을 위해서는 추가적인 전문가 고용 및 시스템 운영에 상당한 비용이 수반하게 된다.
- <9> 따라서, 종래의 신용 위험의 관리를 위한 시스템은 신용 위험에 관한 지식을 갖춘 소수의 전문가에 의해서만 모형 구축이 이루어지고, 시스템의 규모가 상당히 커서 이를 운용하기 위해서는 막대한 자본이 소요되고, 시스템의 유지 및 활용을 위해서는 추가적인 전문가가 필요하고, 이를 위한 상당한 비용이 수반되는 문제점이 있다.

발명의 내용

해결 하고자하는 과제

- <10> 따라서, 본 발명이 해결하고자 하는 첫 번째 과제는 시스템의 규모가 작으면서도 대용량의 신용관리의 데이터를 신속하게 처리 및 유지할 수 있으며, 사용자에게 다양한 신용관리 시스템의 활용도를 부여할 수 있는 신용 위험 모형 구축용 툴 키트를 제공하는 것이다.
- <11> 그리고, 본 발명이 해결하고자 하는 두 번째 과제는, 상기 신용 위험 모형 구축용 툴 키트를 이용하여 신용 위험 관리를 수행할 수 있는 신용 위험 모형 구축을 통한 신용 위험 관리 방법을 제공하는 것이다.
- <12> 또한, 본 발명이 해결하고자 하는 세 번째 과제는, 상기 신용 위험 구축을 통한 신용 위험 관리 방법을 컴퓨터에서 실행시키기 위한 프로그램을 컴퓨터로 읽을 수 있는 기록매체를 제공하는 것이다.

과제 해결수단

- <13> 상기 첫 번째 과제를 해결하기 위하여 본 발명은,
- <14> 신용도 평가를 위하여 기업 또는 개인의 재무제표에 대한 데이터를 저장하는 원자료 데이터베이스, 상기 원자료 데이터베이스로부터 데이터를 선택하여 상기 선택된 데이터의 결측치를 제거하고, 상기 결측치가 제거된 데이터를 관리하는 자료 관리부, 상기 자료 관리부의 데이터를 저장하는 분석용 데이터베이스 저장부, 상기 분석용 데이터베이스 저장부의 데이터로부터 신용도 평가 모형의 변수를 선택하고, 상기 선택된 변수로부터 로지스틱 회귀분석에 의해 데이터로부터 신용도 평가 모형을 구축하는 모형 구축부, 상기 모형 구축부로부터 구축된 신용도 평가 모형의 분석 결과를 저장하는 분석 결과 저장 데이터베이스부, 및 상기 분석 결과 저장 데이터베이스부를 이용하여 상기 기업 또는 개인의 신용도를 평가하는 모형 평가부를 포함하는 신용 위험 모형 구축용 툴 키트를 제공한다.
- <15> 그리고, 상기 두 번째 과제를 해결하기 위하여 본 발명은,
- <16> 기업 또는 개인의 신용도 평가를 위해 산출된 데이터가 저장된 원자료 데이터베이스로부터 분석하고자 하는 데이터를 선택하는 단계, 상기 선택된 데이터에서 상기 신용도 평가에 영향을 주는 변수인 독립변수와 상기 독립변수에 따른 신용도 평가의 결과를 나타내는 변수인 반응변수를 지정하고, 상기 독립변수 및 반응변수를 이용하여 상기 선택된 데이터의 기초통계량 정보를 산출하는 단계, 상기 선택된 데이터 중 상기 독립변수 내에서 결측치를 검출하여 데이터를 정제하는 단계, 상기 정제된 데이터를 기반으로 기초통계량, 상기 독립변수 간의 상관도 및 회귀분석을 수행하여 상기 신용도 평가를 수행할 독립변수를 선별하는 단계, 상기 선별된 독립변수를 이용하여 신용도 평가에 관한 예측 테이블을 구성하고, 로지스틱 회귀분석에 따라 상기 구성된 예측 테이블을 이용하여 신용도 평가 모형을 구축하는 단계, 및 상기 구축된 신용도 평가 모형을 이용하여 상기 신용도 평가값을 산출하여 신용도를 추정하는 단계를 포함하는 신용 위험 모형 구축을 통한 신용 위험 관리 방법을 제공한다.
- <17> 또한, 상기 세 번째 과제를 해결하기 위하여 본 발명은,
- <18> 상기 신용 위험 구축을 통한 신용 위험 관리 방법을 컴퓨터에서 실행시키기 위한 프로그램을 컴퓨터로 읽을 수 있는 기록매체를 제공한다.

효 과

- <19> 본 발명에 의하면, 신용 위험관리의 분석을 위하여 현재 보유하고 있는 데이터를 특수한 형태로 변형할 필요가 없으며, 사용자의 상황에 따라 다양한 유형의 문서 데이터를 이용할 수 있고, 작은 규모의 신용 위험관리를 위한 툴 킷으로 대용량의 데이터를 신속하게 처리할 수 있으며, 이상치 및 결측치 등과 같은 복잡한 작업을 그래픽 유저 인터페이스 환경 하에서 간편하게 처리할 수 있고, 변수 선택 분석방법에 따라 신용 위험관리를 위한 다양한 모형을 제공하며, 현재 많이 사용되는 부도예측 모형에 따른 호환성을 제공하고, 기업의 부도 예측을 위한 효과적인 기업의 분류규칙을 제공할 수 있는 효과가 있다.

발명의 실시를 위한 구체적인 내용

- <20> 이하, 본 발명의 바람직한 실시예를 첨부도면에 의거하여 상세히 설명하기로 한다.
- <21> 그러나, 다음에 예시하는 본 발명의 실시예는 여러 가지 다른 형태로 변형할 수 있으며, 본 발명의 범위가 다음에 상술하는 실시예에 한정되는 것은 아니다. 본 발명의 실시예는 당 업계에서 평균적인 지식을 가진 자에게 본 발명을 더욱 완전하게 설명하기 위하여 제공된다.
- <22> 도 1은 본 발명에 따른 신용 위험 모형 구축용 툴 킷을 블록도로 도시한 것이다.
- <23> 도 1을 참조하면, 본 발명에 따른 신용 위험 모형 구축용 툴 킷(100)은 원자료 데이터베이스(110), 자료 관리부(120), 분석용 데이터베이스 저장부(130), 모형 구축부(140) 및 분석 결과 저장 데이터베이스부(150)로 이루어진다.
- <24> 우선, 원자료 데이터베이스(110)는 신용도 평가를 위하여 기업 또는 개인의 재무제표에 대한 데이터를 저장한다.
- <25> 본 발명에서는 사용자의 작업 단위로 분석 결과를 저장하므로, 이를 수행하기 위하여 별도의 데이터 통합 작업을 할 필요가 없다. 즉, 서로 다른 형식의 다양한 데이터 소스를 하나의 작업 단위에서 불러들일 수 있고, 하나의 작업에서 불러들인 각 데이터 소스는 별도의 데이터 명으로 저장이 가능하다.
- <26> 한편, 본 발명에서는 다양한 형태의 데이터를 이용할 수 있으므로, 일반적으로 쉽게 이용되는 마이크로소프트(Microsoft)사의 엑셀(Excel) 파일이나 다양한 텍스트(text) 형태의 파일을 모두 불러들여 신용도에 대한 분석을 수행할 수 있다.
- <27> 자료 관리부(120)는 상기 원자료 데이터베이스(110)로부터 데이터를 선택하고, 상기 선택된 데이터의 결측치를 제거하며, 결측치가 제거된 데이터를 관리한다.
- <28> 이를 위하여 상기 자료 관리부(120)는 데이터 정의 모듈(121), 데이터 정제 모듈(122), 데이터 대체 모듈(123), 기초 통계 분석 모듈(124), 상관 분석 모듈(125) 및 회귀 분석 모듈(126)로 이루어질 수 있다.
- <29> 상기 데이터 정의 모듈(121)은 상기 원자료 데이터베이스(110)로부터 선택된 데이터에 있어서, 신용도 평가에 영향을 주는 변수인 독립변수와 상기 독립변수에 다른 신용도 평가의 결과를 나타내는 변수인 반응변수를 지정하고, 상기 독립변수 및 반응변수를 이용하여 선택된 데이터의 기초 통계량 정보를 산출할 수 있다.
- <30> 여기서 독립변수는 신용도 평가에 관한 신용 위험 변수(Risk factor)를 의미할 수 있으며, 상기 변수의 지정에 있어서 독립변수 및 반응변수 이외에도 데이터의 시계열 변수(Time), 데이터의 크기(Size) 또는 데이터의 카테고리(Category)를 사용자가 지정할 수 있다.
- <31> 그리고, 상기 데이터 정의 모듈(121)로부터 산출되는 기초 통계량 정보는 원자료 데이터베이스(110)로부터 선택된 데이터의 타입, 각 변수의 역할, 데이터 표본의 크기, 결측치의 개수, 평균, 분산 및 데이터 다섯숫자요약값을 포함할 수 있다.
- <32> 상세히 살펴보면, 기업의 부도예측을 위한 모형 구축을 위한 자료를 이용할 경우, 부도 결과를 나타내는 변수는 반응변수(Response)로 정의할 수 있고, 일반적인 재무제표들은 신용위험변수(Risk factor)로 정의할 수 있다. 대기업 또는 중소기업으로 구분되는 기업 규모 같은 범주형 자료가 존재한다면 이는 카테고리(Category)로 지정할 수 있으며, 년, 분기, 월 또는 일과 같이 시간을 나타내는 변수가 포함되어 있다면 이를 시간(time) 변수로 지정할 수 있다.
- <33> 각 변수의 역할은 상기 데이터의 각 변수의 역할을 지정하여, 상술한 신용 위험 변수(Risk factor), 반응변수

(Response) 이외에도 시간(Time), 크기(Size) 또는 카테고리(Category)를 지정할 수 있다.

- <34> 그리고, 데이터 표본의 크기는 아라비아 숫자로 표기할 수 있으며, 결측치의 개수(NMiss)의 정보를 표시한다.
- <35> 신용도 분석을 위하여 준비된 데이터 중에서는 다른 값들에 비하여 극단적으로 큰 값을 가지거나, 극단적으로 작은 값을 가지는 자료들이 존재할 수 있다. 이러한 데이터를 결측치(missing value)라 하고, 신용도 분석을 수행하는데 있어서 결측치 들을 그대로 사용할 경우, 데이터의 신뢰도에 오류가 발생할 수 있다. 따라서, 실제 신용도 분석을 수행하기 전에 상기 선택된 변수 내에서 결측치라고 파악되는 자료들을 자동적으로 탐색하여 결측치의 총 개수(NMiss)를 출력한다.
- <36> 리고, 상기 기초 통계량 정보에는 평균(Mean), 분산(Variance), 표준편차(Standard Deviation)의 정보를 포함할 수 있으며, 데이터 표본의 평균의 정확도에 대한 기준을 제공하는 표준 오차(Standard error) 값을 포함할 수 있다.
- <37> 아울러, 기초 통계량 정보에는 데이터 다섯숫자요약 값을 포함할 수 있는데, 데이터 다섯숫자요약 값은 최소값(Minimum), 제 1사분위수 값(1st Quartile), 중간값(Median), 제 3사분위수 값(3rd Quartile) 및 최대값(Maximum)을 포함한다.
- <38> 여기서, 제 1사분위수 값(1st Quartile)은 자료를 크기순으로 배열하였을 경우, 하위 25%에 해당하는 값의 범위를 의미하며, 중간값(Median)은 자료를 크기순으로 배열하였을 경우의 중간에 위치하는 값을 의미하며, 자료의 크기인 N이 홀수이면, (N+1/2) 번째 값이 중간값이며, N이 짝수이면 N/2 번째 값과 (N/2)+1 번째 값의 평균이 중간 값에 해당한다.
- <39> 그리고, 제 3사분위수 값(3rd Quartile)은 자료를 크기순으로 정렬하였을 경우 상위 25%의 값을 의미할 수 있다.
- <40> 데이터 정제 모듈(122)은 상기 선택된 데이터를 상기 독립변수 내에서 결측치를 검출하여 데이터를 정제한다.
- <41> 분석을 위하여 준비된 데이터에는 다른 값들에 비하여 극단적으로 큰 값을 가지거나 극단적으로 작은 값을 가지는 자료들이 존재할 수 있다. 이런 결측치(Missing value)들이 데이터 분석에서 그대로 사용된다면 잘못된 결과를 일으킬 수 있으므로 실제 분석에 들어가기 이전에 처리할 수 있는 데이터 정제(Data Cleansing)를 수행한다. 데이터 정제 모듈(122)을 이용하여 결측치 제거 이전과 이후 변수의 속성 변화 과정을 바로 확인할 수 있으며 그 결과는 자동적으로 저장될 수 있다.
- <42> 예를 들어, 우불량 기업 평가에 있어서 데이터의 정제는 데이터의 분포를 안정화시키고 정제된 데이터를 통해 신용 위험에 대한 판별력을 제고하기 위하여 결측치를 제거하는 것이다. 그리고, 데이터에 대하여 사용자가 선택한 변수 혹은 전체에 대하여 기능을 수행하고 수정된 데이터에 대한 처리 결과를 사용자에게 제공할 수 있다.
- <43> 데이터 대체 모듈(123)은 상기 데이터 정제 모듈(122)에서 검출된 결측치를 대체 데이터로 대체한다.
- <44> 즉, 데이터 대체 모듈(123)은 신용도 평가 모형에 사용하고자 하는 자료가 결측치를 가지고 있는 경우 사용자가 미리 설정한 데이터로 대체하여 전체 분석용 데이터에 자료를 추가하는 역할을 수행한다.
- <45> 한편, 신용도 평가 모형을 구축하는데 있어서 분석을 수행하고자 하는 자료에서 상기 결측치의 비율이 너무 많은 경우 이는 정보의 손실을 의미하게 된다. 특히 결측치의 발생이 구조적인 차원에서 이루어진 것이라면 결측치 자체가 충분한 정보를 포함하고 있으므로 이를 신용도 평가 모형에 이용하여야 한다.
- <46> 신용도 평가 모형에 상기 결측치를 이용하기 위해서는 상기 결측치를 다른 값으로 대체를 수행하여야 하는데, 본 발명에서는 이를 위하여 평균 대체 방법으로 관찰된 값들의 평균을 연산하여 결측치를 상기 연산된 평균값으로 대체하거나, 중위수 대체 방법으로 관찰된 값들의 중간값을 연산하여 상기 결측치를 연산된 중간값으로 대체할 수 있다. 상기 대체된 결측치는 새로운 데이터의 형태로 상기 데이터 정제 모듈(122)에 저장될 수 있다.
- <47> 그리고, 기초 통계 분석 모듈(124)은 상기 신용 위험의 관리 모형을 구축하는데 필요한 각 독립변수들에 대한 기초통계량을 연산한다. 기초통계량은 상기 데이터 정의 모듈(121)에서 각 변수들이 정의되면 사용자에게 의해 연산될 수 있고, 데이터 정제 모듈(122)에서 데이터가 정제된 후 기초통계량이 연산될 수 있다.
- <48> 상관 분석 모듈(125)은 신용도 평가 모형을 구축하는데 필요한 각 독립변수들 간의 상관계수를 연산한다. 상기 신용도 평가 모형의 구축을 위하여 위험요인(Risk factor)으로 사용되는 각 독립변수들, 재무제표 간의 상관계수를 연산하고, 연산된 상관계수에 대한 통계적 유의성을 검증하고 검증된 유의성 결과를 바로 조회할 수 있다.

이를 통하여 실제 신용도 평가 모형의 구축을 하기 전 각종 재무제표들 간의 연관성을 검증하고, 이를 실제로 분석할 수 있도록 지원한다.

- <49> 회귀분석 모듈(126)은 상기 반응변수가 다른 독립변수들에 의하여 어떻게 설명되어지는지 또는 어떻게 예측되는지를 알아보기 위하여 독립변수인 위험요인이 반응변수에 미치는 영향력의 정도를 비교하여, 상기 독립변수의 종속변수에 대한 영향력에 대한 유의성 검증을 수행한다. 이는 실제 금융 분야에서 시장모형, 자본시장선 모형과 증권 시장선 모형과 같은 자본자산 가격결정 모형 등에 적용될 수 있다.
- <50> 분석용 데이터베이스 저장부(130)는 상기 회귀분석 모듈(126)에 의해 통계적 유의성 검증을 거친 데이터를 저장한다.
- <51> 상기 분석용 데이터베이스 저장부(130)에 저장된 데이터는 모형 구축부(140)에서 신용도 평가 모형의 변수를 선택하고, 상기 선택된 변수로부터 로지스틱 회귀분석 또는 역회귀분석에 의해 데이터로부터 신용도 평가 모형을 구축한다.
- <52> 한편, 모형 구축부(140)는 세그멘테이션(segmentation) 모듈(141), 독립 변수 선택 모듈(142), 로지스틱 모형 모듈(143) 및 캘리브레이션(calibration) 모듈(144)을 포함할 수 있다.
- <53> 세그멘테이션 모듈(141)은 독자적인 판별 모형을 이용한 분류규칙을 제공한다. 이는 기업이나 개인의 각종 재무제표를 이용하여 미리 정의된 그룹(segment)에 따라 개별 기업의 분류를 수행할 수 있는 적합한 모형을 구축하고, 상기 구축된 모형을 바탕으로 한 분류규칙을 제공한다. 구축된 분류규칙에 의하여 개별 기업에 대한 그룹을 결정할 수 있고, 또한 아직 그룹이 정해지지 않은 새로운 개인이나 기업에 대한 분류 예측을 수행할 수 있게 한다. 이와 같은 기능을 적용함으로써 개별 개인이나 기업에 대한 우량 또는 불량으로 구분을 수행할 수 있고, 금융기관에서는 이와 같은 개인 또는 기업의 데이터를 이용하여 각 신용상태에 따른 우량 또는 불량여부를 판별하고, 보험회사 등에서는 고객의 이탈 방지를 위한 세그멘테이션(segmentation) 구축이나 고객 등급의 부여 등에 활용된다.
- <54> 독립 변수 선택 모듈(142)은 기업이나 개인을 대상으로 하는 신용 위험 관리 모형을 구축하는데 있어서 경우에 따라 매우 방대한 요인들에 사용하여 신용도 평가 모형을 구축할 수 있는데 너무 많은 위험요인, 즉 독립변수를 하나의 모형에 모두 사용하는 것은 분석의 효율을 떨어뜨릴 수 있으며, 구축된 모형의 활용도를 악화시킬 수 있다.
- <55> 본 발명에서는 신용도 평가 모형 구축의 효율성과 구축된 모형의 활용도를 높이기 위하여 적절한 위험요인의 선택 기준을 제공한다. 신용도 평가의 모형을 구축하는데 빈번하게 사용되는 일반화 선택 모형을 기반으로 하여 각종 위험요인 선택을 위한 통계량을 제시함으로써 사용자는 본 발명에서 제공되는 기초 통계량을 확인하여 적절한 위험요인 즉, 독립변수를 선택할 수 있다.
- <56> 로지스틱 모형 모듈(143)은 개인이나 기업의 신용도 평가 모형을 구축하는데 있어서, 개인이나 기업의 신용도 예측을 위하여 신용 위험에 영향을 미치는 주요한 위험요인을 확인할 수 있으며, 구축된 모형에 대한 통계적 유의성 및 안정성을 기초 통계량을 통하여 확인할 수 있다. 그리고, 상기 위험요인에 따른 독립변수를 이용하여 신용도 평가를 위한 컷 오프 포인트(cut off point)를 추출하여 신용도를 예측하게 된다.
- <57> 캘리브레이션 모듈(144)은 상기 로지스틱 모형 모듈을 이용하여 신용도를 추정할 경우 기업 또는 개인의 재무제표에 대한 데이터에 근거하여 신용도를 평가함으로써, 실제 기업이나 개인의 신용도와 일정한 차이를 보일 수 있다. 이와 같은 차이를 실제 기업이나 개인의 신용도에 맞게 재조정하기 위하여 신용도의 재추정을 캘리브레이션 모듈을 이용하여 수행한다.
- <58> 상기 신용도의 재추정을 수행하기 위하여 비모수 회귀 모형을 이용한 역회귀 방법을 이용하고, 특히 부분 가중 회귀법(Locally Weighted Robust Regression Method: LOESS)을 이용하여 신용도 재추정을 수행할 수 있다. 이와 같은 재추정 과정을 통하여 실제 신용도에 근접한 신용도 평가값을 추정하게 된다.
- <59> 이와 같은 신용도 평가값은 분석 결과 저장 데이터베이스부(150)에 저장되어 사용자가 필요시마다 상기 신용도 평가값을 활용할 수 있게 된다.
- <60> 이와 더불어, 상기 원자료 데이터베이스(110), 분석용 데이터베이스 저장부(130) 및 분석 결과 저장 데이터베이스부(150)에 저장된 데이터는 XML 문서 형식으로 저장될 수 있으므로 웹상에서 필요에 따라 용이하게 사용자가 사용할 수 있다.

- <61> 도 2a는 본 발명에 적용되는 원자료 데이터베이스에서 선택된 데이터를 도시한 것이다.
- <62> 도 2a를 참조하면 우선, 상술한 바와 같이 원자료 데이터베이스에 저장된 데이터를 사용자의 선택에 의해 불러올 수 있다.
- <63> 사용자의 선택에 의해 불러온 데이터 시트의 화면에서 상단의 첫 번째 행은 데이터의 각 변수를 의미하며, 각 열은 각 변수에 따른 실제 데이터를 표시한 것이다.
- <64> 도 2b는 선택된 데이터에서 데이터의 변수 정보를 도시한 것이다.
- <65> 도 2b를 참조하면, 'Role'에 해당되는 영역이 변수의 역할을 의미하며, 각 변수마다 변수의 역할을 지정할 수 있다. 이는 상술한 바와 같이 독립변수와 독립변수에 대응되는 결과값을 의미하는 반응변수인 종속변수를 지정할 수 있다.
- <66> 도 2b에서 독립변수는 'Risk Factor'를 의미하고, 종속변수인 반응변수는 'Response'로 지정되고, 그 외에 시간인 'Time', 데이터의 크기인 'Size', 데이터의 카테고리 'Category'등을 사용자에게 의해 지정할 수 있으며, 상기 데이터의 변수 정보를 지정함으로써 실질적인 신용도 평가를 수행할 수 있게 된다.
- <67> 도 2c는 선택된 변수에 의해 출력된 기초통계량 정보를 도시한 것이다.
- <68> 이는 상기 도 2b에서 지정한 변수에 따라 선택된 데이터로부터 선택된 데이터의 타입(TYPE), 각 변수의 역할(Role), 데이터 표본의 크기(N), 결측치의 개수(NMiss), 평균(Mean), 분산(Variance) 표준편차(Standard) 및 데이터 다섯숫자요약 값을 포함할 수 있다.
- <69> 도 3a는 본 발명에 적용되는 데이터 정제 화면을 도시한 것이고, 도 3b는 도 3a에 의해 선택된 데이터의 데이터 정제 결과를 도시한 것이다
- <70> 우선 도 3a를 참조하면, 선택된 데이터 시트에서 각 독립변수 별로 데이터의 정제 여부를 결정한다. 여기서 데이터를 정제할 독립변수는 'Cleaning'을 선택하고, 정제하지 않을 데이터에 대한 독립변수는 'None'을 선택한다.
- <71> 그러면, 상기 선택된 결과에 따라 도 3b에서 볼 수 있는 바와 같이 데이터 정제 결과를 확인할 수 있다.
- <72> 도 3b를 참조하면, 데이터 정제 전(Uncleaned)에는 결측치에 해당되는 데이터가 없었으나, 데이터 정제 후(cleaned)에는 13개의 결측치가 늘어난 것을 확인할 수 있다.
- <73> 이와 더불어 상자 그림(Box plot)을 통하여 데이터 정제 후의 데이터의 분포 양상을 그래프로 확인할 수 있다.
- <74> 상기 상자 그림에서 라인(311)은 선택된 데이터의 중간값을 나타내고, 라인(312)은 선택된 데이터의 중간값에서 하위 25%의 값에 해당되는 데이터의 최소값을 표시한 것이고, 라인(313)은 선택된 데이터의 중간값으로 부터 상위 25%의 값에 해당되는 데이터의 최대값을 표시한 것이다.
- <75> 여기서, 선택된 데이터의 중간값에서 하위 25%의 값에 해당되는 데이터의 최소값인 라인(312)의 값과 선택된 데이터의 중간값으로 부터 상위 25%의 값에 해당되는 데이터의 최대값인 라인(313) 사이의 범위를 사분위 범위(Inter Quartile Range:IQR)(310)라 지칭한다.
- <76> 그리고, 상기 선택된 데이터의 중간값에서 하위 25%의 값에 해당되는 데이터의 최소값인 라인(312)으로부터 하위로 상기 $IQR \times 1.5$ 의 값에 해당하는 라인을 최소 한계값 라인(320)으로 지칭하고, 상기 선택된 데이터의 중간값에서 상위 25%의 값에 해당되는 데이터의 최대값인 라인(312)으로부터 상위로 상기 $IQR \times 1.5$ 의 값에 해당하는 라인을 최대 한계값 라인(330)으로 지칭한다.
- <77> 그러면, 상기 최소 한계값 라인(320)과 최대 한계값 라인(330)의 범위를 벗어나는 값을 결측치라고 하고, 결측치는 다른 값들에 비하여 극단적으로 큰 값을 가지거나 극단적으로 작은 값을 가지기 때문에 이러한 결측치들이 그대로 사용될 경우 잘못된 결과를 일으킬 수 있다.
- <78> 따라서, 이러한 결측치 데이터를 정제할 독립변수에 'Cleaning'을 선택한 후 결측치가 탐지되었을 경우, 상기 결측치를 제외한 기초통계량 정보를 상기 도 3b와 같이 확인함으로써 보다 신뢰성 있는 기초통계량 정보를 얻을 수 있다.
- <79> 도 4a는 도 3b의 결측치를 대체하는 것을 도시한 것이다.
- <80> 도 4a를 참조하면, 도 3b에서 결측치가 존재할 경우, 상기 결측치가 존재하는 변수에 대하여 상기 결측치를 이

용하기 위해서는 상기 결측치를 다른 값으로 대체를 수행하여야 하는데, 본 발명에서는 이를 위하여 평균 대체 방법으로 관찰된 값들의 평균을 연산하여 상기 결측치를 평균값(Mean imputation)으로 대체하거나, 중위수 대체 방법(Median imputation)으로 관찰된 값들의 중간값을 연산하여 상기 결측치를 대체할 수 있다. 상기 대체된 결측치는 새로운 데이터의 형태인 평균값 또는 중간값으로 저장되게 된다.

- <81> 도 4b는 도 4a에 의해 결측치가 대체된 데이터의 정보를 도시한 것이다.
- <82> 도 4b를 참조하면, 도 4a에서 평균값(Mean imputation)으로 관찰된 값들의 평균값을 연산하여 X4의 독립변수에 해당하는 상기 결측치를 대체한 경우, 13개의 결측치 값이 모두 평균값으로 대체되어 결측치가 존재하지 않음을 확인할 수 있다. 아울러, 데이터의 분산된 정도를 나타내는 분산의 값이 146.94에서 127.45로 감소하여 데이터가 편차가 줄어들었음을 알 수 있고, 데이터 다섯숫자요약 값의 폭 역시 줄어들어 해당되는 값들의 퍼짐 정도가 감소하였다.
- <83> 도 5는 본 발명에 적용되는 데이터의 상관분석도를 도시한 것이다.
- <84> 본 발명은, 데이터의 상관분석을 통하여 신용도 위험 관리를 위하여 적절한 변수를 탐색하는 과정을 수행할 수 있다.
- <85> 우선, 사용자에게 의해 데이터의 상관분석을 수행할 독립변수를 복수 개를 선택한 후 상기 독립변수 간 상관분석을 수행하면 신용도 위험 관리에 적절한 독립변수를 선정할 수 있다.
- <86> 아울러, 상관 분석표와 함께 정제된 데이터의 기초통계량 정보(510)를 확인할 수 있으며, 상관계수의 통계적 유의성(520) 검증을 동시에 수행할 수 있다.
- <87> 도 5를 참조하면, 5개의 독립변수가 상관분석에 사용된 경우를 도시한 것으로, 가장 먼저 5개의 독립변수에 해당하는 데이터의 기초 통계량 정보를 확인할 수 있고, 이와 더불어 행렬로 표시된 상관분석도를 확인할 수 있다.
- <88> 행렬로 표시된 상기 상관분석도에는 사용자가 미리 정의한 유의수준을 정의할 수 있다. 본 실시예에서는 유의수준을 0.05로 정의하였고, 유의수준 0.05이하에서 유의한 경우는 하얀 색으로 p-value가 도시됨을 확인할 수 있다.
- <89> 상기 상관분석의 결과에서 주의하여 보아야 할 사항은 변수 간 상관계수 가운데 그 계수가 가장 큰 값이다. 도 5를 참조하면 첫 번째 변수인 X1(510)이 나타내는 총자본 경상이익률과 두 번째 변수인 X3(530)이 나타내는 자기자본 경상이익률 간의 상관계수이다. 두 변수 간 상관계수는 매우 밀접한 연관관계를 가지고 있다고 할 수 있으며, 만약 부도예측모형을 위한 로지스틱 회귀모형에서 이 두 변수가 모두 독립변수로 사용되는 경우 다중공선성이라는 문제가 발생할 수 있다.
- <90> 일반적으로 독립변수 간 상관계수가 0.7 이상이 되면 다중공선성을 의심할 수 있으며, 상관계수가 0.9 이상이 되면 심각하게 다중공선성을 의심할 수 있다. 다중공선성이 발생하게 되면 정상적인 회귀분석을 수행할 수 없게 되며 개별 회귀계수에 대한 추론에서 잘못된 추론결과를 도출하는 오류를 범할 수 있다.
- <91> 이는 도 5에서 볼 수 있는 바와 같이 X1(510)과 X3(530)의 상관계수가 0.9를 초과하였으므로 실제 모형구축을 위하여 로지스틱 회귀분석을 수행하고자 할 경우, 두 변수 X1(510)과 X3(530) 가운데에서 하나의 변수만을 독립변수(Risk factor)로 사용하여야 한다.
- <92> 도 6은 본 발명에 적용되는 로지스틱 회귀분석에 의한 독립변수 선택을 위한 분석결과를 도시한 것이다.
- <93> 도 6을 참조하면, 신용도 평가에 사용될 후보 변수인 독립변수를 하나씩 선택하여 그 결과를 출력하는 것을 알 수 있다. 이는 로지스틱 회귀모듈에 의해 산출된다.
- <94> 로지스틱 회귀모듈에 따라 선택의 기준으로 사용되는 독립변수의 통계량은 아카이케 정보 기준(Akaike Information Criterion:AIC), 슈바르츠 베이저안 기준(Schwarz Bayesian Criterion:SBC) 및 정확도 비율(Accuracy Ratio)이 있다. 이 가운데에서 AIC와 SBC는 그 값이 작을수록 신용도 분석에 적합한 독립변수가 될 수 있고, 정확도 비율은 그 값이 클수록 신용도 분석에 적합하다.
- <95> 도 6에서 Observation(610)은 회귀모듈에 사용된 회사의 부도여부를 평가하기 위한 각 독립변수별 총 자료의 수를 의미한다.
- <96> Missing(620)은 회귀모듈에 사용된 각 독립변수별 결측치에 해당되는 자료의 수를 의미한다.

<97> Default Count(630)는 부도여부의 예측을 위한 회귀모형에서 반응변수인 부도여부 변수에서 실제 부도로 나타난 데이터의 수를 의미한다. 분석에서 사용되는 예제자료에서 전체 99개의 회사 가운데에서 2개의 회사가 부도 처리된 회사임을 나타낸다.

<98> Estimate(640)는 부도확률을 p 라 하였을 때, 각 독립변수(Risk Factor)에 대한 회귀분석에서 하기의 수학적 식 1에 사용되는 회귀계수의 추정치인 β_1 을 의미한다. 여기서, β_0 는 하기 수학적 식 1에서 절편에 해당하는 값으로 입력될 독립변수(Risk Factor)가 없을 경우에 적용되는 값을 의미한다.

수학적 식 1

$$\text{logit}(p) = \log \frac{p}{1-p} = \beta_0 + \beta_1 x$$

<99>

<100> 그리고, 상술한 아카이케 정보 기준(Akaike Information Criterion:AIC)(650)은 적합한 모형들 간에 신용 모형 적합도를 비교하기 위하여 사용되는 하나의 척도로써 이 값이 적을수록 신용도 모형의 적합도가 우수하고, 이를 선호되는 모형이라고 할 수 있다. 여기서 AIC는 하기의 수학적 식 2와 같이 연산될 수 있다.

수학적 식 2

$$AIC = -2 \log \sum_j f_j \log(p_j) + 2(k + s)$$

<101>

<102> 수학적 식 2에서 f_j 는 j 번째의 관찰치의 빈도를 의미하고, p_j 는 j 번째 관찰치의 추정된 부도확률, k 는 반응변수의 범주-1 으로 여기서는 1을 의미할 수 있고, S 는 독립변수의 수를 의미하고, 상기의 단순 로지스틱 회귀모형에서는 1을 나타낸다.

<103> 또한, 슈바르츠 베이지안 기준(Schwarz Bayesian Criterion:SBC)(660)은 상기의 AIC와 마찬가지로 적합한 모형들 간에 모형 적합도를 비교하기 위하여 사용되는 하나의 척도로써, 이 값이 작을수록 모형의 적합도가 우수하여 선호되는 모형이라 할 수 있다. 한편, 상기 슈바르츠 베이지안 기준인 SBC 값은 하기의 수학적 식 3과 같이 계산될 수 있다.

수학적 식 3

$$SBC = -2 \log \sum_j f_j \log(p_j) + (k + s) \log(\sum_j f_j)$$

<104>

<105> 마찬가지로, 수학적 식 3에서 f_j 는 j 번째의 관찰치의 빈도를 의미하고, p_j 는 j 번째 관찰치의 추정된 부도확률, k 는 반응변수의 범주-1 으로 여기서는 1을 의미할 수 있고, S 는 독립변수의 수를 의미하고, 상기의 단순 로지스틱 회귀모형에서는 1을 나타낸다.

<106> 또한, 정확도 비율(Accuracy Ratio)(670)은 반응변수의 값(여기서는 부도여부)와 모형에 의해 예측된 부도 확률 간의 측정된 값으로써 이 값이 클수록 모형의 적합이 더 우수하다고 할 수 있다.

<107> 상기 정확도 비율의 측정방법은 하기의 수학적 식 4와 같다.

수학적 식 4

$$\text{Accuracy Ratio} = \frac{n_c - n_d}{t}$$

<108>

<109> 상기 수학적 식 4를 기반으로 한 정확도 비율의 측정방법은 전체 자료 가운데에서 2개의 자료를 추출하고, 이 경우

두 개의 값은 서로 다른 반응변수를 가져야 한다. 즉, 한 변수의 값이 부도라면 다른 한 변수의 값이 부도가 아니어야 한다. 이렇게 추출된 두 개의 자료를 하나의 쌍으로 고려한다. 여기서 가능한 모든 쌍의 값을 t 라고 정의하면, 각 쌍에서 모형에 의해 계산된 부도 확률을 확인한다. 이 경우 각 쌍에서 부도가 아닌 회사와 부도인 회사의 예측된 부도확률이 같은 방향이라면, 즉, 부도인 회사의 부도 확률이 부도가 아닌 회사의 부도 확률보다 크다면 이를 일치 쌍이라고 하고, 같은 방향이 아니라면, 즉, 부도인 회사의 부도확률이 부도가 아닌 회사의 부도확률보다 작다면, 이를 불일치 쌍이라 한다. 일치쌍과 불일치쌍의 전체 수를 각각 n_c 와 n_d 라 하고 하여 상기의 수식식 4와 같이 정확도 비율이 측정될 수 있다.

- <110> 도 7a는 본 발명의 로지스틱 회귀모형을 이용한 신용도 평가 모형을 구축할 경우의 독립변수와 반응변수를 도시한 것이다.
- <111> 실질적으로 신용도 평가 모형에 있어서 최종적으로 알아내고자 하는 것은 신용도, 예를 들면 기업의 부도 여부이다.
- <112> 특정 기업에 대하여 그 기업의 각종 재무제표를 고려할 경우, 그 회사가 부도처리가 될 것인지($y=0$), 부도가 되지 않을 것인지($y=1$)의 여부를 통계적 모형으로 판별하고자 하는 것이다. 로지스틱 회귀분석은 이와 같이 반응변수인 y 가 2개의 값을 가지는 자료에 대하여 통계적 모형으로 적합한 분석방법이라 할 수 있다.
- <113> 로지스틱 회귀 모형을 이용하여 특정 기업의 부도 여부에 유의미한 영향을 미치는 독립 변수를 확인할 수 있고, 각 독립 변수 간의 상대적인 비교를 통하여 부도 여부에 미치는 영향력의 정도를 비교할 수 있다.
- <114> 그리고, 모형의 최종 목적인 각 기업별 예측 부도확률을 추정할 수 있으며, 적합한 모형의 평가를 수행하고 부도 여부에 대한 적절한 컷 오프 포인트를 찾기 위한 테이블을 함께 제공할 수 있다.
- <115> 예를 들어, 개인이나 기업의 원자료 데이터베이스 가운데 3개의 변수를 선택하여 부도예측모형을 로지스틱 회귀분석의 모형을 이용하여 모형 적합도를 수행한다.
- <116> 이는 자료의 특성에 따라 모형이 적합이 이루어지면, 그 결과들은 모두 결과화면에서 확인할 수 있다. 각 결과로부터 구축된 로지스틱 회귀모형의 적합성, 모형 적합과 관련된 각종 기초통계량, 모형의 실행을 확인할 수 있는 테이블, 신용도를 판정하기 위한 컷 오프 포인트를 탐색하기 위한 테이블 및 최종적으로 분석에 사용된 모든 기업 자료에 대한 추정된 기업별 부도예측 확률을 확인할 수 있다.
- <117> 우선, 도 7a를 참조하면, 로지스틱 회귀 분석(Logistic Regression Analysis) 창에서는 분석을 위하여 사용자 지정을 수행할 수 있다.
- <118> 첫 번째로, 사용자는 하나의 프로젝트에서 수행할 수 있는 복수의 데이터 셋(Data Set)을 지정할 수 있다. 만약, 데이터를 선택하는 과정에서 복수 개의 데이터셋을 호출할 경우, 호출된 데이터 셋의 명칭이 출력된다. 이 가운데 사용자는 모형 구축을 위한 데이터 셋을 선택(Select Data Set)(711)할 수 있고, 만약 프로젝트 내에서 호출된 데이터 셋이 하나일 경우 별도의 지정이 필요 없다.
- <119> 두 번째로, 하나의 프로젝트에서 복수 개의 모형 구축이 가능한데, 이 경우 구축된 모형을 별도로 지정할 수 있고, 이를 위하여 모형의 명칭을 지정(Input Name)(712)한다.
- <120> 세 번째로, 구축하고자 하는 로지스틱 회귀모형의 반응변수를 지정(Target Variable)(713)한다. 예를 들면, 기업의 부도예측에 관한 모형을 구축하고자 하는 경우, 회사의 부도 여부를 나타내는 변수를 지정한다. 이는 데이터를 정의할 때 반응변수를 지정하는 것으로 수행될 수 있다.
- <121> 네 번째로, 구축하고자 하는 로지스틱 회귀모형의 독립변수를 지정(Independent Variable)(714)한다. 기업의 부도예측 모형을 구축하고자 하는 경우에는 각 재무제표를 독립변수로 고려하여 지정할 수 있고, 이는 그래픽 유저 인터페이스 환경에서 지정 가능하다.
- <122> 다섯 번째로, 모형구축에 사용된 독립변수에 대한 기초통계량 및 상관분석 결과를 출력할 경우에 기초통계량(Descriptive Statistics)(715)을 선택함으로써 상기 기초통계량 정보를 출력할 수 있다.
- <123> 여섯 번째로, 구축하고자 하는 로지스틱 회귀모형에 인터셉트(intercept)를 추가하고자 하는 경우 인터셉트 항(Intercept)(716)을 지정할 수 있다. 만약 이 지정을 하지 않을 경우, 상기 인터셉트 항이 없는 로지스틱 회귀모형을 이용하여 모형 적합 여부를 판별한다.

- <124> 상기 여섯 개의 항목을 지정한 다음, 출력되는 분석 결과는 총 5개의 영역으로 나눌 수 있다.
- <125> 첫 번째로, 모델 정보(Model Information)로 모형의 추정과정에서 오류가 발생하거나 정상적으로 수렴이 이루어졌는지 확인할 수 있다. 정상적으로 모형 추정이 이루어진 경우 특별한 오류 메시지 없이 '반복 완료(Iteration finished)' 메시지를 확인할 수 있고, 만약 추정과정에서 문제가 발생할 경우 자동적으로 오류 메시지가 출력된다.
- <126> 이는 도 7b에서 볼 수 있는 바와 같이 로지스틱 회귀모형의 모형 수렴 결과를 확인할 수 있다.
- <127> 두 번째로, 파라미터 추정(Parameter Estimate)으로, 이는 적합된 로지스틱 회귀모형에서의 각 회귀계수에 대한 추정치와 관련된 값들이 출력된다. 이 경우 각 독립변수의 유의성 여부도 함께 확인할 수 있다.
- <128> 상세히 살펴보면, 부도확률 예측 모형 구축을 위한 로지스틱 회귀모형에서 반응변수인 Y는 기업별 부도여부를 나타낸다. 만약 기업이 부도이면, $y=1$ 이고, 정상이면 $y=0$ 이 된다. 그리고 주어진 독립변수의 값에 따라 개별 기업의 부도 확률을 $p = \Pr(y=1|X=x)$ 라 할 때 모형 구축을 위한 로지스틱 회귀모형은 하기의 수학적 식 5와 같다.

수학적 식 5

$$\text{logit}(p) = \log \frac{p}{1-p} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$$

- <129>
- <130> 수학적 식 5로부터 각 회귀계수 β 들에 대한 추정치는 뉴턴-랩슨(Newton-Raphson) 방법을 이용하여 하기의 수학적 식 6에서 연달아 연산되는 $\beta^{(r+1)}$ 과 $\beta^{(r)}$ 가 거의 일치할 때까지 반복 수행하여 수렴하는 값이 각 회귀계수에 대한 추정치가 된다.

수학적 식 6

$$\text{logit}(p) = \log \frac{p}{1-p} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$$

- <131>
- <132> 세 번째로, 모델 적합성 정보(Model Fit Information)으로 적합된 로지스틱 회귀모형이 된 원자료에 대한 적합도를 확인하기 위한 각종 통계량을 제공한다.
- <133> 이는 도 7c에서 볼 수 있는 바와 같이 본 발명의 로지스틱 회귀모형의 위험요인, 즉 독립변수에 대한 통계량 정보를 확인할 수 있다. 즉, 제공된 통계량을 이용하여 적합된 모형을 이용하여 부도 예측 등에 적용하는데 있어서, 모형에 대한 독립변수의 적합 정도를 확인할 수 있게 된다.
- <134> 본 정보는 -2loglikelihood, AIC, SBC, chi-square, p-value 및 정확도 비율 등의 정보를 포함한다.
- <135> 이 가운데 AIC와 SBC는 다른 독립변수를 사용하여 구축된 모형과 상대적인 적합도를 비교하는데 사용된다. 이 값은 작을수록 모형 적합이 잘 이루어진다고 할 수 있다.
- <136> chi-square과 p-value는 적합된 로지스틱 회귀모형에서 모든 회귀계수가 0이다 라는 귀무가설을 검증하는데 사용된다. 정확도 비율은 적합된 결과가 실제 부도여부를 얼마나 잘 설명하는가에 대한 측도로써 1과 0 사이의 값을 가지며, 이 값이 클수록 모형 적합도가 우수하다고 할 수 있다.
- <137> 이는 도 7d에 도시된 로지스틱 회귀분석에 따른 모델 적합도를 살펴보면, 모델 적합성 여부에 대한 귀무가설 $H_0: \beta_1 = \beta_2 = \beta_3 = 0$ 에 검정결과 p-value가 0.01보다 더 작으므로 귀무가설을 기각할 수 있다. 즉, 적합된 모형이 실제 자료를 적절히 설명하고 있다고 할 수 있다. 이는 도 7d를 살펴보면, 정확도 비율(Accuracy Ratio)을 살펴보았을 때 0.9355로 1에 가까우므로 모형 적합이 적절하다고 할 수 있다.
- <138> 한편, -2loglikelihood 통계량은 현재 추정된 모형의 로그 우도 함수에 -2를 곱한 값으로 하기의 수학적 식 7과 같이 연산될 수 있다.

수학식 7

$$-2\log\text{likelihood} = -2\sum_j f_j \log(p_j)$$

상기 수학식 7에서 f_j 는 j 번째의 관찰치의 빈도를 의미하고, p_j 는 j 번째 관찰치의 추정된 부도확률을 나타낸다.

그리고, chi-square는 추정된 회귀계수를 가지고 모집단 회귀계수에 대한 검정인 $H_0: \beta=0$ 를 위한 검정통계량을 나타낸다. 로지스틱 회귀분석에서 개별 회귀계수들에 대한 검정은 왈드 검정(wald test)을 사용한다. 이 때 검정통계량은 이미 연산된 추정치와 추정치의 표본오차를 이용하여 하기의 수학식 8과 같이 연산될 수 있으며, 귀무가설하에서 자유도가 1인 카이제곱 분포를 따른다. 자유도가 1인 카이제곱 분포로부터 연산된 검정통계량을 이용하여 p-value를 연산할 수 있다.

수학식 8

$$\chi^2 = (\beta / s_\beta)^2 \sim \chi^2(1)$$

네 번째로, 분류 테이블(Classification Table)은 정 분류 그래프를 작성하기 위한 수치들을 표의 형태로 제공하여 주는 테이블이다. 즉, 분류 테이블과 정 분류 그래프는 상호 보완적으로 사용될 수 있다.

분류 테이블(Classification Table)은 적합된 부도확률 예측 모형의 결과를 이용하여 예측 부도확률을 위한 적절한 컷 오프 포인트를 탐색하고, 적합된 모형이 실제 부도확률을 예측하는데 있어서 얼마나 정확한 결과를 도출해 내는가를 평가하는데 사용되는 각종 수치들을 제공하는 테이블이다. 그리고 정 분류 그래프는 이 분류 테이블을 한 눈에 파악할 수 있도록 그래프로 도시한 것이다.

본 발명에 따른 분류 테이블은 7e에 도시된 바와 같다. 도 7e를 참조하면, 연산된 부도예측 확률로부터 실제 부도 여부를 예측하는 컷 오프 포인트의 확률을 0.005 단위로 제공하고 있다. 예를 들어 도 7e에서 컷 오프 포인트를 0.020으로 할 경우, 모형 적합 결과 연산된 부도확률을 가지고 부도확률이 0.02 이상이면 부도로 판정하고, 그렇지 않을 경우 정상으로 판정함을 의미한다.

판정결과와 하기의 표 1의 분류 테이블(Classification table)과 같이 정리할 수 있다.

표 1

구 분		예 측		합 계
		정상	부도	
실제	정상	87	6	93
	부도	1	1	2
합 계		88	7	95

상기 표 1의 판정결과 분류 테이블에 의하면 실제 부도인 기업을 부도로 예측한 경우는 1 건으로 분류 테이블에서 정확한 이벤트(Correct Event)에 해당한다.

또한, 판정결과 실제 정상인 기업을 정상으로 예측한 경우는 87건으로 분류 테이블에서 정확한 비이벤트(Correct Non-Event)에 해당한다.

한편, 판정결과 실제 부도인 기업을 정상으로 잘못 예측한 경우는 1건으로 분류 테이블에서 오류 비이벤트(Incorrect Non-Event)에 해당한다.

그리고, 판정결과 실제 정상인 기업을 부도로 잘못 예측한 경우는 6건으로 분류 테이블에서 오류 이벤트(Incorrect Event)에 해당한다.

- <152> 상기 분류 테이블을 기반으로 하여 다시 하기와 같이 확률을 계산할 수 있다.
- <153> 첫 번째로 정확 예측률(Percentage Correct)로 이는 정상을 정상으로 부도를 부도로 정확하게 예측할 확률이다.
- <154> 두 번째로는 민감도(Sensitivity)로 실제 부도 기업을 부도로 예측할 확률이다.
- <155> 세 번째로는 특이도(Specificity)로 실제 정상인 기업을 정상으로 예측할 확률이다.
- <156> 상기 표 1을 기반으로 정확 예측률, 민감도 및 특이도를 정리하면 하기의 표 2와 같다.

표 2

구분	수식	분포도수	확률
정확 예측률	$\frac{Freq(실제=부도\&예측=부도)+Freq(실제=정상\&예측=정상)}{Freq(Total)}$	$\frac{87+1}{95}$	92.6%
민감도	$\frac{Freq(실제=부도\&예측=부도)}{Freq(실제=부도)}$	$\frac{87}{93}$	50%
특이도	$\frac{Freq(실제=정상\&예측=정상)}{Freq(실제=정상)}$	$\frac{87}{93}$	93.5%

- <157>
- <158> 상기 표 1 및 표 2를 기반으로 산출된 각종 빈도와 상기 빈도로부터 산출된 3가지의 확률인 정확 예측률(Percentage Correct), 민감도(Sensitivity) 및 특이도(Specificity)가 0.005 단위로 증가하는 컷 오프 포인트 별로 모두 연산되며 상기 3가지의 확률이 정분류 그래프(Correct Classification Plot)를 통하여 제공된다. 상기 정분류 그래프는 도 7f에 도시된 바와 같다. 도 7f를 참조하면, 도 7e에 의한 분류 테이블을 활용하여 최적의 컷 오프 포인트를 구할 수 있다.
- <159> 도 7g는 상기 도 7f의 정분류 그래프를 기반으로 생성한 예측 테이블을 도시한 것이다.
- <160> 도 7g를 참조하면, 예측 테이블(Prediction table)은 상기 데이터 셋의 모든 개별 기업에 대하여 구축된 모형으로부터 연산된 예측 부도 확률과 실제 부도 여부를 하나의 테이블로 정리한 것이다.
- <161> 이는 분석에 사용된 99개의 기업 데이터 가운데 일부인 20개 기업에 대한 실제 부도여부와 예측부도확률을 나타낸 테이블이다. 도 7g에서 식별자(ID)가 10번에 해당하는 기업의 실제 값이 1로 부도 기업을 나타낸다. 그리고, 이 기업에 대한 모형으로부터 예측된 부도 확률(Predict)은 0.8841임을 볼 수 있으므로 매우 근접한 부도 확률이 연산되었음을 알 수 있다.
- <162> 한편, 식별자가 16인 기업의 경우, 기업의 실제 값이 0으로 정상인 기업이나, 예측된 부도 확률은 0.114466으로 다른 예측값보다 상대적으로 큰 것을 확인할 수 있다.
- <163> 이는 상기 분류 테이블에서 살펴본 바와 같이 만약 컷 오프 포인트를 0.02로 설정한다면 0.114466은 0.02를 훨씬 상회하는 값으로 부도 기업으로 잘못 예측되는 것이다.
- <164> 개별기업에 대한 예측확률의 연산은 추정된 로지스틱 회귀계수를 이용하여 수행한다. 로지스틱 회귀모형에서의 추정치를 $\beta_i, i=1, \dots, k$ 라 할 때, 상기 추정치를 이용하여 적합된 로지스틱 회귀모형은 상기 수학적 식 6과 같다.
- <165> 상기 수학적 식 6으로부터 각 개별 기업에 대한 예측 부도확률 p는 하기의 수학적 식 9에 의해 연산된다.

수학적 식 9

$$p = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k)}$$

- <166>
- <167> 도 8a는 본 발명의 캘리브레이션 모듈에 의해 역회귀분석을 수행하는 데이터 셋 지정화면을 도시한 것이다.
- <168> 도 8a에서 볼 수 있는 바와 같이 캘리브레이션 모듈에 의해 역회귀 분석을 수행하기 위한 데이터 셋과 모델명을 지정한다. 한편, 분석 시 다양한 조건을 바탕으로 다양한 모델을 만들었을 경우, 캘리브레이션을 수행할 각각의

모델에 대한 지정을 수행할 수 있다.

- <169> 이와 더불어 분할할 구간 개수에 대한 지정을 할 수 있다. 이는 나누어진 구간에 따라 전체적인 예측부도확률을 재추정할 것이고, 이를 통하여 보다 현실적인 그래프를 출력한다. 여기서 나누어진 구간의 최소값은 20일 수 있고, 최대값은 50일 수 있다.
- <170> 도 8b는 본 발명의 캘리브레이션 모듈에 의해 역회귀분석의 옵션 지정 화면을 도시한 것이다.
- <171> 도 8b를 참조하면, 이는 비모수적 회귀분석 방법에 대하여 지정할 수 있는 옵션을 도시한 것을 확인할 수 있다. 비모수적 회귀분석 방법(Non-linear Regression Method)이나 전술한 LOESS 방법을 선택할 수 있다.
- <172> LOESS방법은 다양한 옵션을 선택할 수 있는데 독립변수의 차원(Order of the polynomial), 가중치의 적용 방법(Function used to determine the weight), 분석에 필요한 반복 횟수(Number of iteration), 평활 모수(smoothing parameter)를 지정할 수 있으며, 평활 모수의 값은 0.20에서 0.78까지 다양하게 지정할 수 있다. 상기 분석에 필요한 반복 횟수는 1-5까지 지정 가능하다.
- <173> 도 8c는 본 발명의 역회귀분석의 사용 데이터를 도시한 것이다.
- <174> 도 8c를 참조하면, 분석 수행시 선택한 옵션에서 구간을 20개로 나눌 경우, 독립변수인 X1의 값은 1부터 20까지 입력될 것이고, 반응변수인 y값은 로지스틱 회귀분석을 통하여 추정된 예측 부도 확률을 20구간으로 나누고 그 구간의 실제 부도확률을 나타내게 된다.
- <175> 도 8d는 상기 도 8c에 따른 캘리브레이션 결과를 출력한 그래프이다.
- <176> 도 8d를 참조하면, 부도여부의 예측 모형에 의한 예측확률과 LOESS에 의한 적합 확률을 동시에 도시하고 있음을 알 수 있다.
- <177> LOESS 방법을 적용한 결과 기존 예측 모형에 의한 예측확률보다 보다 평활(Smooth)한 형태, 즉 전체적으로 단조 증가하는 함수의 형태로 재표현되고 있음을 알 수 있다.
- <178> 도 9는 본 발명의 신용 위험 모형 구축을 통한 신용 위험 관리 방법의 흐름도를 도시한 것이다.
- <179> 도 9를 참조하면, 우선, 기업 또는 개인의 신용도 평가를 위해 산출된 데이터가 저장된 원자료 데이터베이스로부터 분석하고자 하는 데이터를 선택한다(910 과정).
- <180> 즉, 기업 또는 개인의 신용도 평가를 위하여 산출된 데이터가 저장된 데이터베이스로부터 사용자는 직접 평가하고자 하는 데이터를 선택하는데, 본 발명에서는 다양한 형태의 데이터를 이용할 수 있으므로, 일반적으로 쉽게 이용되는 엑셀 파일이나 다양한 텍스트 형태의 파일을 모두 불러들여 신용도에 대한 분석을 수행할 수 있다.
- <181> 상기 선택된 데이터에서 상기 신용도 평가에 영향을 주는 변수인 독립변수와 상기 독립변수에 따른 신용도 평가의 결과를 나타내는 변수인 반응변수를 지정하고, 상기 독립변수 및 반응변수를 이용하여 상기 임포트(import)된 데이터의 기초통계량 정보를 산출한다(920 과정).
- <182> 이는 원자료 데이터베이스로부터 선택된 데이터에서 신용도 평가 영향을 주는 독립변수와 상기 독립변수에 따른 신용도 평가의 결과를 나타내는 변수인 반응변수를 지정한다.
- <183> 여기서 독립변수는 신용도 평가에 관한 신용 위험 변수를 의미할 수 있고, 상기 변수의 지정에 있어서 데이터의 시계열 변수, 데이터의 크기 또는 데이터의 카테고리를 사용자가 지정할 수도 있다.
- <184> 그리고, 기초통계량 정보는 원자료 데이터베이스로부터 선택된 데이터의 타입, 각 변수의 역할, 데이터 표본의 크기, 결측치의 개수, 평균, 분산 및 상술한 데이터의 다섯숫자요약 값을 포함할 수 있다.
- <185> 그 다음, 상기 선택된 데이터 중 상기 독립변수 내에서 결측치를 검출하여 데이터를 정제한다(930 과정).
- <186> 이는, 우불량 기업의 평가에 있어서 선택된 자료에 결측치를 포함하고 있는 경우, 사용자가 미리 설정한 데이터로 대체하여 전체 분석용 데이터에 자료를 추가하는 역할을 수행하게 된다.
- <187> 본 발명에서는 이를 위하여 평균 대체 방법으로 결측치로 관찰된 값들을 선택된 자료의 평균값으로 대체하거나, 중위수 대체 방법으로 관찰된 값들의 중간값을 연산하여 결측치를 대체한다.
- <188> 상기 정제된 데이터를 기반으로 기초통계량, 상기 독립변수 간의 상관도 및 회귀분석을 수행하여 상기 신용도

평가를 수행할 독립변수를 선별한다(940 과정).

- <189> 상기 기초 통계량은 정제된 데이터를 기반으로 연산된 값이므로 이 값을 통하여 데이터 정제 전과 데이터 정제 후의 선택된 데이터의 분포 상태 등을 알 수 있고, 신용도 평가 모형을 구축하는데 필요한 각 독립변수의 상관도를 분석하여 신용도 평가 모형을 구축하는데 사용되는 각 독립변수인 재무제표들 간의 상관계수를 연산하여, 연산된 상관계수에 대한 통계적 유의성을 검증하고, 검증된 유의성 결과를 조회한다.
- <190> 그런 다음, 상기 반응변수가 다른 독립변수에 의하여 어떻게 설명되어지는지를 알아보기 위하여 회귀 분석을 수행하여 독립변수인 위험요인이 반응변수에 미치는 영향력의 정도를 비교하여, 상기 독립변수의 종속변수에 대한 영향력의 유의성 검증을 수행한다. 이는 실제 금융 분야에서 시장 모형, 자본 시장선 모형과 증권 시장선 모형과 같은 자본 자산 가격결정 모형 등에 적용될 수 있다.
- <191> 그 다음, 상기 선별된 독립변수를 이용하여 신용도 평가에 관한 예측 테이블을 구성하고, 상기 구성된 예측 테이블을 이용하여 신용도 평가 모형을 구축한다(950 과정).
- <192> 이는, 신용도 평가 모형을 구축하는데 있어서, 로지스틱 회귀분석을 수행하거나, 상기 로지스틱 회귀분석 후 역 회귀분석을 함께 수행하여 신용도 평가 모형을 구축할 수 있다.
- <193> 신용도 평가 모형은 상기 독립변수를 이용하여 신용도 평가에 관한 적절한 컷 오프 포인트를 탐색하고, 적합한 모형이 실제의 신용도를 예측하는데 있어서 얼마나 정확한 결과를 도출해 내는지의 여부를 평가하는데 사용되는 각종 수치들을 제공하는 예측 테이블에 의해 신용도 평가 모형을 구축할 수 있다.
- <194> 한편, 신용도 평가 모형은 로지스틱 평가 모형을 포함할 수 있다.
- <195> 로지스틱 평가 모형에서 사용자는 하나의 프로젝트에서 수행할 수 있는 복수의 데이터 셋을 지정할 수 있는데, 데이터를 선택하는 과정에서 복수 개의 데이터 셋을 호출할 수 있다. 그러면, 호출된 데이터 셋을 기반으로 신용도의 모형을 지정하고, 모형의 명칭을 지정할 수도 있다.
- <196> 그런 후, 구축하고자 하는 로지스틱 회귀모형의 반응변수, 독립변수를 지정하여 모형의 추정과정에서 오류가 발생하였는지 정상적으로 수립이 이루어졌는지에 대한 모델 정보, 파라미터 추정, 적합한 로지스틱 회귀모형이 된 원자료에 대한 적합도를 확인하기 위한 각종 통계량을 제공하는 모델 적합성 정보, 정 분류 그래프를 작성하기 위한 수치들을 표의 형태로 제공하여 주는 분류 테이블을 로지스틱 평가 모형은 제공한다.
- <197> 마지막으로, 상기 구축된 신용도 평가 모형을 이용하여 상기 신용도 평가값을 산출하여 신용도를 추정한다(960 과정).
- <198> 즉, 이는 상기 구축된 신용도 평가 모형에 의해 산출된 컷 오프 포인트에 따라서 신용도의 우불량 여부를 추정할 수 있다.
- <199> 그리고, 본 발명은 비모수 회귀모형을 이용하여 신용도 평가값을 재추정하고, 재추정된 신용도 평가값을 이용하여 신용도를 평가할 수 있다.
- <200> 이는 상기 로지스틱 모형을 이용하여 신용도를 추정할 경우 기업 또는 개인의 재무제표에 대한 데이터에 근거하여 신용도를 평가함으로써, 실제 기업이나 개인의 신용도와 일정한 차이를 보일 수 있다. 이와 같은 차이를 실제 기업이나 개인의 신용도에 맞게 재조정하기 위하여 신용도의 재추정을 수행한다.
- <201> 상기 신용도의 재추정을 수행하기 위하여 비모수 회귀 모형을 이용한 역회귀 방법을 이용할 수 있고, 특히 부분가중 회귀법(Locally Weighted Robust Regression Method:LOESS)을 이용하여 신용도 재추정을 수행할 수 있다. 이와 같은 재추정 과정을 통하여 실제 신용도에 근접한 신용도 평가값을 추정하게 된다.
- <202> 이와 같은 신용도 평가값은 분석 결과 저장 데이터베이스에 저장되어 사용자가 필요시 마다 상기 신용도 평가값을 활용할 수 있게 된다.
- <203> 본 발명은 컴퓨터로 읽을 수 있는 기록 매체에 컴퓨터(정보 처리 기능을 갖는 장치를 모두 포함한다)가 읽을 수 있는 코드로서 구현하는 것이 가능하다.
- <204> 컴퓨터가 읽을 수 있는 기록매체는 컴퓨터 시스템에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록 장치를 포함한다. 컴퓨터가 읽을 수 있는 기록 장치의 예로는 ROM, RAM, CD-ROM, 자기 테이프, 플로피 디스크, 광 데이터 저장장치 등이 있다. 또한 컴퓨터가 읽을 수 있는 기록매체는 네트워크로 연결된 컴퓨터 장치에 분산되어 분산방식으로 컴퓨터가 읽을 수 있는 코드가 저장되고 실행될 수 있다.

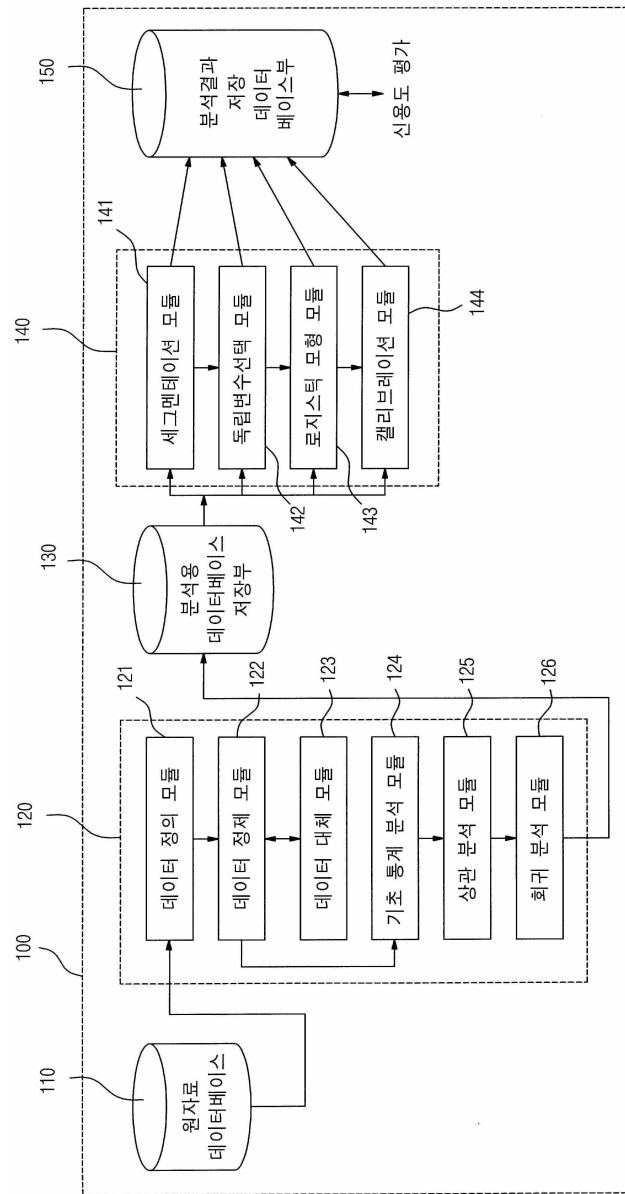
<205> 본 발명은 도면에 도시된 실시예를 참고로 설명되었으나 이는 예시적인 것에 불과하며, 본 기술 분야의 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 타 실시예가 가능하다는 점을 이해할 것이다. 따라서, 본 발명의 진정한 기술적 보호 범위는 첨부된 등록청구범위의 기술적 사항에 의해 정해져야 할 것이다.

도면의 간단한 설명

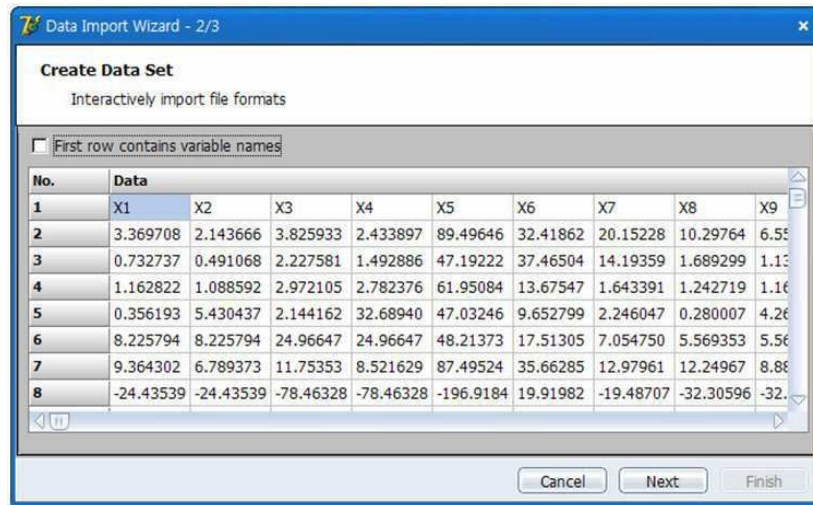
- <206> 도 1은 본 발명에 따른 신용 위험 모형 구축용 툴 킷의 블록도이다.
- <207> 도 2a는 본 발명에 적용되는 원자료 데이터베이스에서 선택된 데이터를 도시한 것이다.
- <208> 도 2b는 선택된 데이터에서 데이터의 변수 정보를 도시한 것이다.
- <209> 도 2c는 선택된 변수에 의해 출력된 기초통계량 정보를 도시한 것이다.
- <210> 도 3a는 본 발명에 적용되는 데이터 정제 화면을 도시한 것이다.
- <211> 도 3b는 도 3a에 의해 선택된 데이터의 데이터 정제 결과를 도시한 것이다.
- <212> 도 4a는 도 3b의 결측치를 대체하는 것을 도시한 것이다.
- <213> 도 4b는 도 4a에 의해 결측치가 대체된 데이터의 정보를 도시한 것이다.
- <214> 도 5는 본 발명에 적용되는 데이터의 상관분석도를 도시한 것이다.
- <215> 도 6은 본 발명에 적용되는 본 발명에 적용되는 로지스틱 회귀분석에 의한 독립변수 선택을 위한 분석결과를 도시한 것이다.
- <216> 도 7a는 본 발명의 로지스틱 회귀모형을 이용한 신용도 평가 모형의 독립변수와 반응변수를 도시한 것이다.
- <217> 도 7b는 본 발명의 로지스틱 회귀모형의 수렴 결과를 도시한 것이다.
- <218> 도 7c는 본 발명의 로지스틱 회귀모형의 독립변수에 대한 통계량 정보를 도시한 것이다.
- <219> 도 7d는 본 발명의 로지스틱 회귀분석의 모델 적합도를 도시한 것이다.
- <220> 도 7e는 본 발명에 적용되는 분류 테이블을 도시한 것이다.
- <221> 도 7f는 도 7e에 의한 컷 오프 포인트를 포함하는 정분류 그래프를 도시한 것이다.
- <222> 도 7g는 도 7f의 정분류 그래프를 기반으로 생성된 예측 테이블을 도시한 것이다.
- <223> 도 8a는 본 발명의 캘리브레이션 모듈에 의해 역회귀분석을 수행하는 데이터 셋 지정화면을 도시한 것이다.
- <224> 도 8b는 본 발명의 캘리브레이션 모듈에 의해 역회귀분석의 옵션 지정 화면을 도시한 것이다.
- <225> 도 8c는 본 발명의 역회귀분석의 사용 데이터를 도시한 것이다.
- <226> 도 8d는 도 8c에 따른 캘리브레이션 결과 출력 그래프이다.
- <227> 도 9는 본 발명의 신용 위험 구축을 통한 신용 위험 관리 방법의 흐름도를 도시한 것이다.

도면

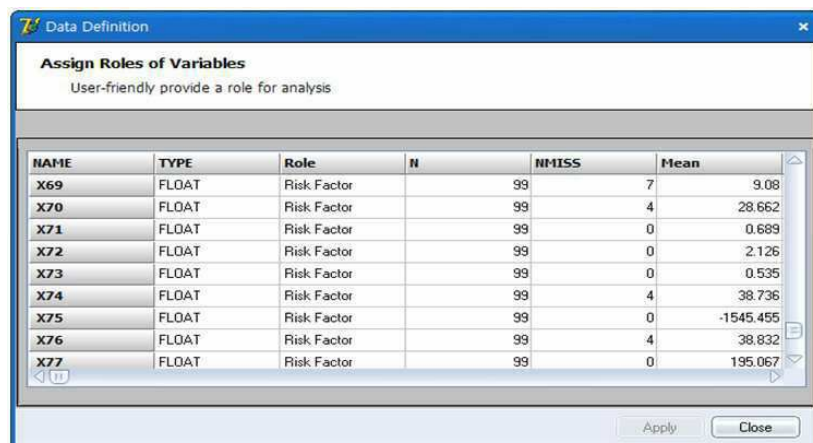
도면1



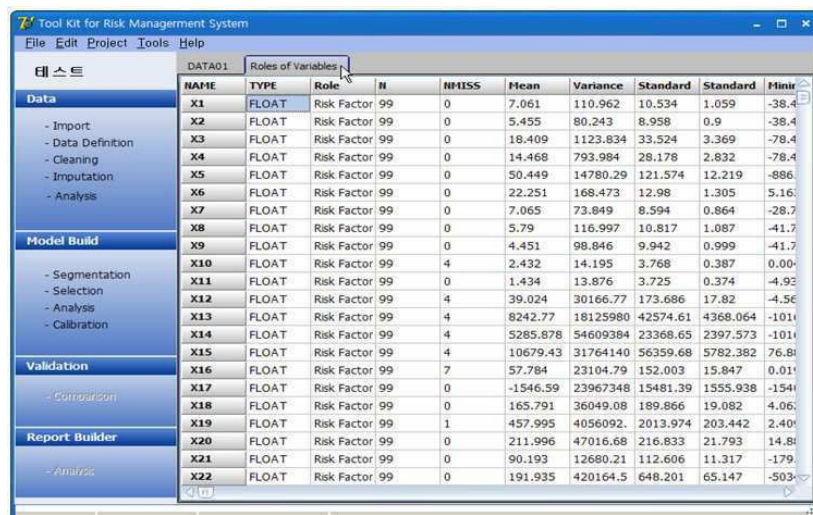
도면2a



도면2b



도면2c



도면3a

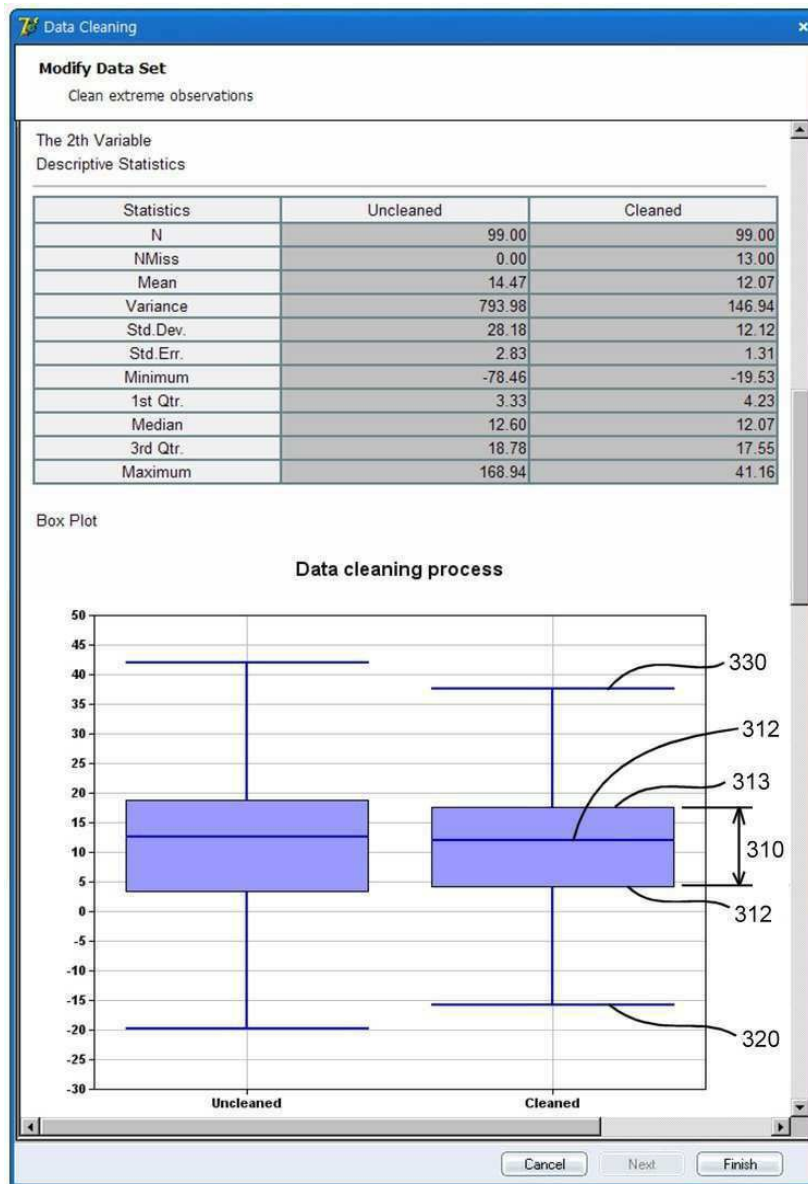
Data Cleaning

Modify Data Set
Clean extreme observations

NAME	Name	Role	Process	Min	Max	Mean	Std.Err	Standard	Stan
X1	FLOAT	Risk Factor	Cleaning	99	0	7.061	110.962	10.534	
X2	FLOAT	Risk Factor	None	99	0	5.455	80.243	8.958	
X3	FLOAT	Risk Factor	None	99	0	18.409	1123.834	33.524	
X4	FLOAT	Risk Factor	None	99	0	14.468	793.984	28.178	
X5	FLOAT	Risk Factor	Cleaning	99	0	50.449	14780.294	121.574	1
X6	FLOAT	Risk Factor	None	99	0	22.251	168.473	12.98	
X7	FLOAT	Risk Factor	None	99	0	7.065	73.849	8.594	
X8	FLOAT	Risk Factor	None	99	0	5.79	116.997	10.817	
X9	FLOAT	Risk Factor	None	99	0	4.451	98.846	9.942	

Cancel Next Finish

도면3b



도면4a

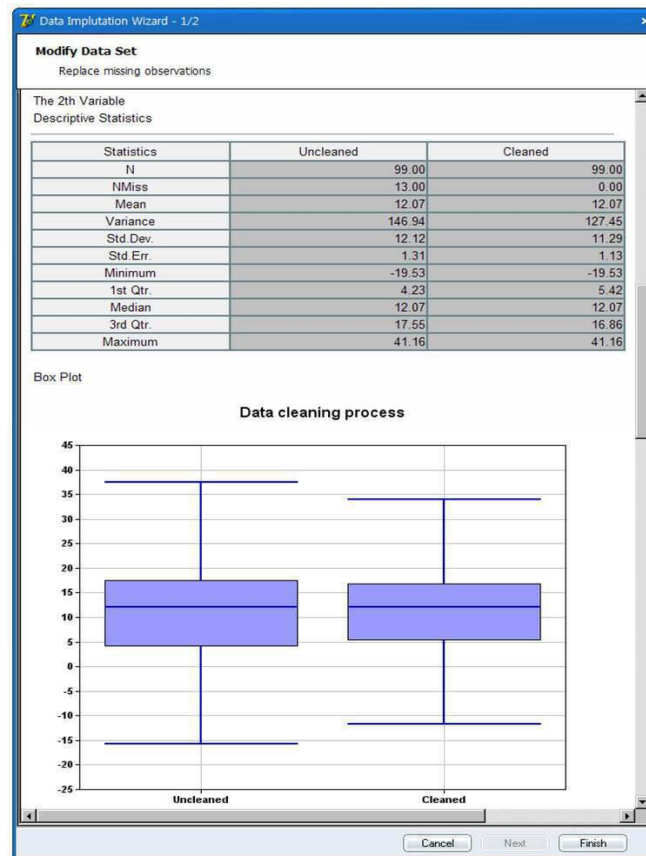
Data Imputation Wizard - 1/2

Modify Data Set
Replace missing observations

NAME	Name	Role	Process	Min	Max	Mean	Std.Err	Stand
X1	FLOAT	Risk Factor	Mean imputation	99	6	6.64	41.195	6
X2	FLOAT	Risk Factor	None	99	0	5.455	80.243	8
X3	FLOAT	Risk Factor	None	99	0	18.409	1123.834	33
X4	FLOAT	Risk Factor	Median imputation	99	13	12.07	146.942	12
X5	FLOAT	Risk Factor	Mean imputation	99	7	66.659	740.299	27
X6	FLOAT	Risk Factor	None	99	0	22.251	168.473	1
X7	FLOAT	Risk Factor	Median imputation	99	4	7.073	35.193	5
X8	FLOAT	Risk Factor	None	99	0	5.79	116.997	10
X9	FLOAT	Risk Factor	Median imputation	99	0	4.451	98.846	9

Cancel Next Finish

도면4b



도면5

Analysis finished at 2007/ 3/ 6 14:21:12

510

Descriptive Statistics

N	99	99	99	99	99
NMiss	0	0	0	4	4
Mean	6.6402	18.4090	12.0702	8242.7695	38.8320
Variance	38.6727	1123.8342	127.4498	1812598018.6409	30171.2821
Std. Dev.	6.2187	33.5236	11.2894	42574.6171	173.6988
Std. Err.	0.6250	3.3693	1.1346	4368.0635	17.8211
Minimum	-6.5553	-78.4633	-19.5268	-1016.9457	-3.9692
1st Qtr.	1.7851	4.2252	5.4200	257.0056	1.0808
Median	6.6402	13.8145	12.0702	673.2475	3.4637
3rd Qtr.	10.4623	24.9665	16.8644	2360.0886	13.9233
Maximum	24.6955	189.8662	41.1559	369734.9451	1552.9451

520

Correlation

Corr	1.0000	0.3826	0.6203	0.0809	0.1573
p-value	0.0000	0.0000	0.0000	0.4156	0.0985
Corr	0.3826	1.0000	0.3772	0.0159	0.0457
p-value	0.0000	0.0000	0.0000	0.8775	0.6518
Corr	0.6203	0.3772	1.0000	0.0461	0.0500
p-value	0.0000	0.0000	0.0000	0.6492	0.6210
Corr	0.0809	0.0159	0.0461	1.0000	0.9479
p-value	0.4156	0.8775	0.6492	0.0000	0.0000
Corr	0.1573	0.0457	0.0500	0.9479	1.0000
p-value	0.0985	0.6518	0.6210	0.0000	0.0000

도면6

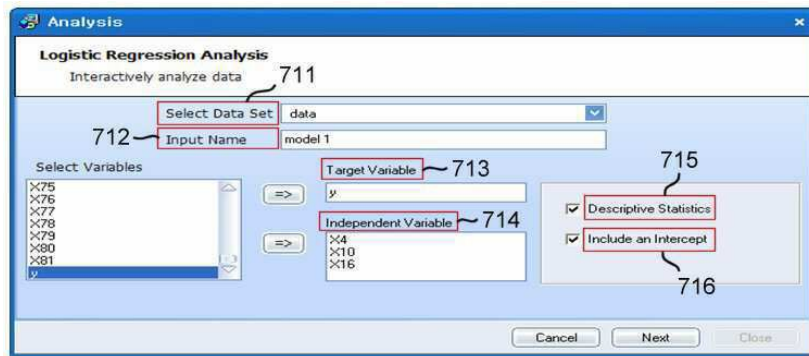
Simple Logistic Regression Models

Variable	Observation	Missing	Default Count	Estimates	AIC	SBC	Accuracy Ratio
1-th	99	0	2	-0.1391	15.7625	20.9528	0.7216
2-th	99	0	2	-0.1440	15.6866	20.8768	0.7629
3-th	99	0	2	0.0272	18.2078	23.3980	0.8041
4-th	99	0	2	0.0387	16.4860	21.6762	0.7320
5-th	99	0	2	0.0015	23.5212	28.7114	0.1134

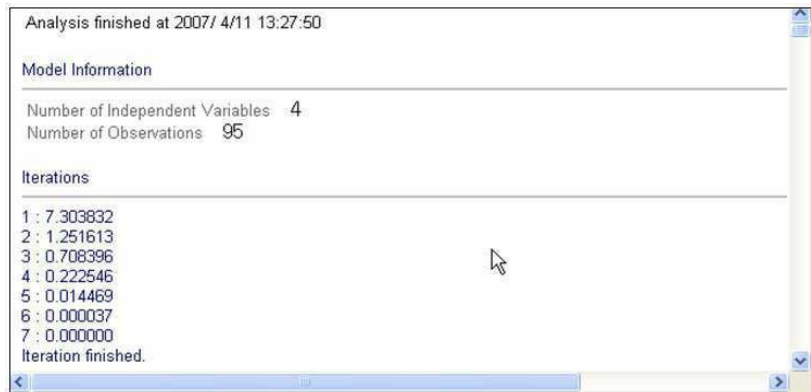
610 620 630 640 650 660 670

SELECTION RESULT: 3th Variable

도면7a



도면7b



도면7c

Risk Factor Selection
Automatic Risk Factor Selection
Using Simple Logistic Regression Models

Page2

Descriptive Statistics

	99	99	99
N	99	99	99
NMiss	0	0	0
Mean	7.0615	5.4550	18.4090
Variance	110.9618	80.2427	1123.8342
Std. Dev.	10.5338	8.9578	33.5236
Std. Err.	1.0587	0.9003	3.3693
Minimum	-38.4563	-38.4563	-78.4633
1st Qtr.	1.2120	1.1438	4.2252
Median	6.2063	5.4304	13.8145
3rd Qtr.	10.7136	9.0044	24.9665
Maximum	57.2626	44.7903	189.8662

Buttons: Cancel, Next, Close

도면7d

Analysis		
Logistic Regression Analysis		
Interactively analyze data		
Model Fit Information		
-2 Loglikelihood		17.2649
AIC		25.2649
SBC		35.4804
Chi-square		54.1155
P-value		0.0000
Accuracy Ratio		0.9355

도면7e

Analysis

Logistic Regression Analysis

Interactively analyze data

Classification Table

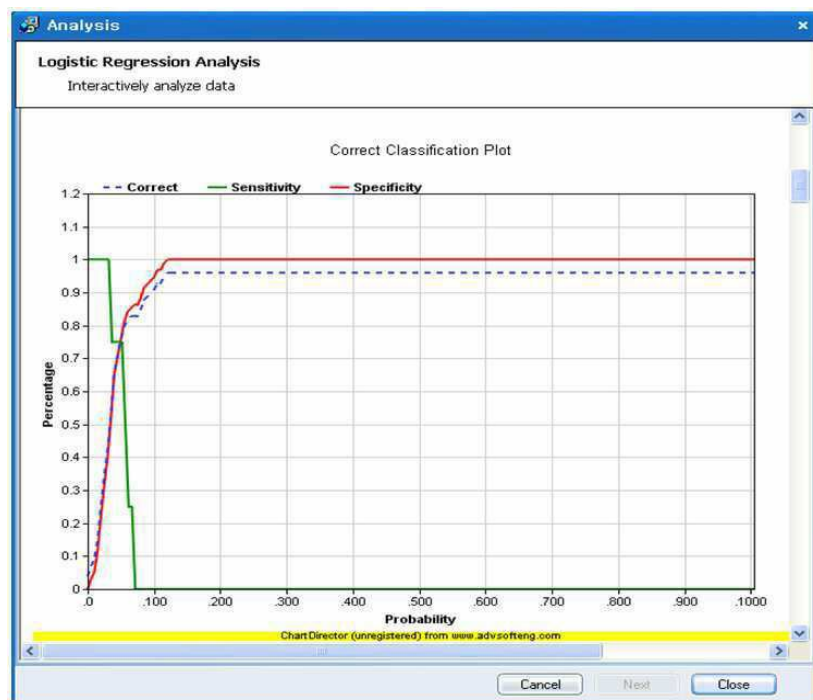
Probability Level	Correct		Incorrect		Percentage		
	Event	Non-Event	Event	Non-Event	Correct	Sensitivity	Specificity
0.000	4	0	95	0	0.040	1.000	0.000
0.005	4	3	92	0	0.071	1.000	0.032
0.010	4	5	90	0	0.091	1.000	0.063
0.015	4	11	84	0	0.152	1.000	0.116
0.020	4	21	74	0	0.253	1.000	0.221
0.025	4	30	65	0	0.343	1.000	0.316
0.030	4	38	57	0	0.424	1.000	0.400
0.035	3	49	46	1	0.525	0.750	0.516
0.040	3	62	33	1	0.657	0.750	0.653
0.045	3	67	28	1	0.707	0.750	0.705
0.050	3	72	23	1	0.758	0.750	0.758
0.055	2	77	18	2	0.798	0.500	0.811
0.060	1	80	15	3	0.818	0.250	0.842
0.065	1	81	14	3	0.828	0.250	0.853
0.070	0	82	13	4	0.828	0.000	0.853
0.075	0	82	13	4	0.828	0.000	0.853
0.080	0	84	11	4	0.848	0.000	0.884
0.085	0	87	8	4	0.879	0.000	0.916
0.090	0	88	7	4	0.889	0.000	0.926
0.095	0	89	6	4	0.899	0.000	0.937

Cancel

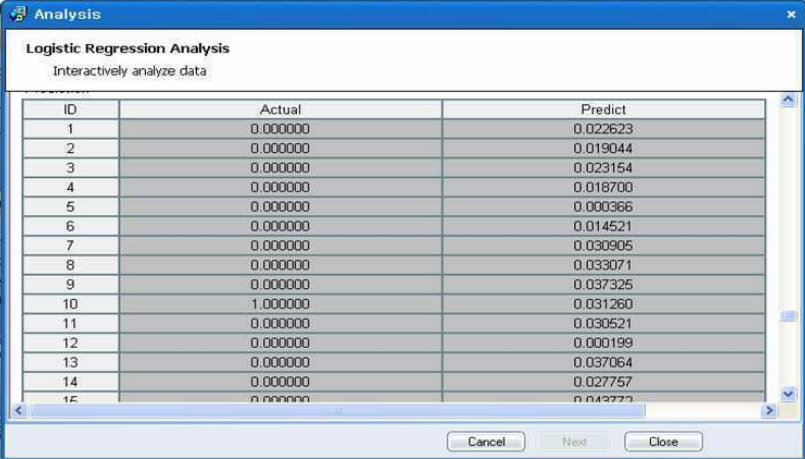
Next

Close

도면7f

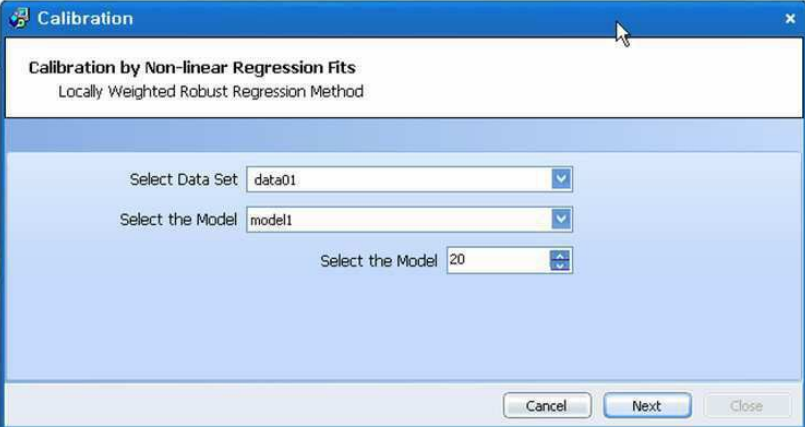


도면7g



ID	Actual	Predict
1	0.000000	0.022623
2	0.000000	0.019044
3	0.000000	0.023154
4	0.000000	0.018700
5	0.000000	0.000366
6	0.000000	0.014521
7	0.000000	0.030905
8	0.000000	0.033071
9	0.000000	0.037325
10	1.000000	0.031260
11	0.000000	0.030521
12	0.000000	0.000199
13	0.000000	0.037064
14	0.000000	0.027757
15	0.000000	0.043773

도면8a



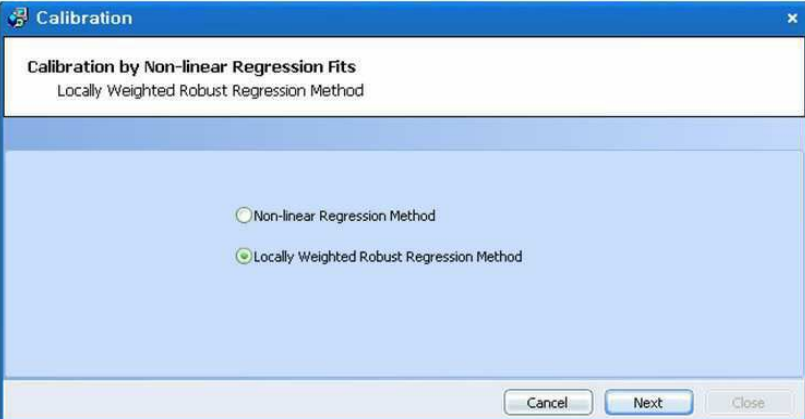
Calibration by Non-linear Regression Fits
Locally Weighted Robust Regression Method

Select Data Set: data01

Select the Model: model1

Select the Model: 20

도면8b



Calibration by Non-linear Regression Fits
Locally Weighted Robust Regression Method

☐ Non-linear Regression Method

☒ Locally Weighted Robust Regression Method

도면8c

Calibration

Calibration by Non-linear Regression Fits
Locally Weighted Robust Regression Method

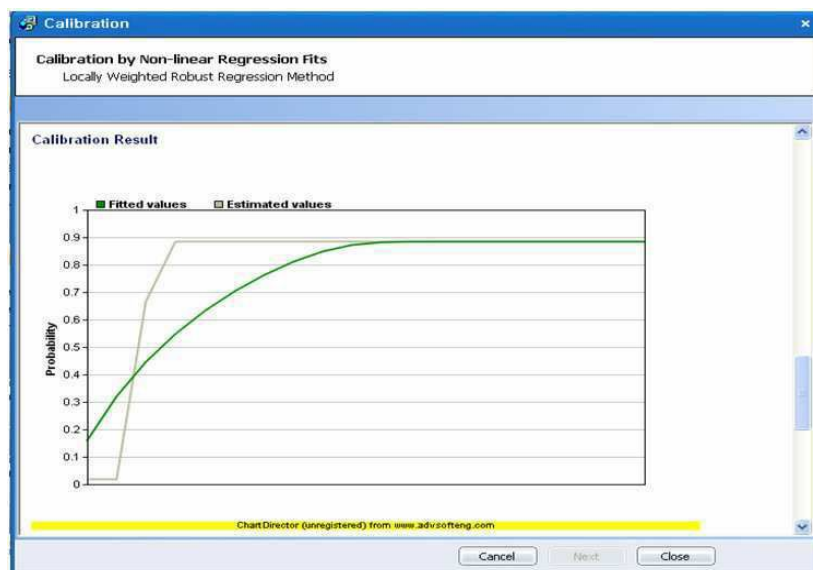
Distribution of Default Data

Model Data

x0	x1	y
1.00	1.00	0.01604
1.00	2.00	0.01604
1.00	3.00	0.66763
1.00	4.00	0.88416
1.00	5.00	0.88416
1.00	6.00	0.88416
1.00	7.00	0.88416
1.00	8.00	0.88416
1.00	9.00	0.88416
1.00	10.00	0.88416

Cancel Next Close

도면8d



도면9

