

이화여자대학교 대학원

2002학년도

석사학위 청구논문

소기업 신용평가 모형 구축에 관한 연구

經 營 學 科

李 宜 洙

2003

소기업 신용평가 모형 구축에 관한 연구

이 論文을 碩士學位 論文으로 提出함

2003年 1月

梨花女子大學校 大學院

經營學科 李 宜 洙

李 宜 洙의 碩士學位 論文으로 認准함

指導教授 申 璟 植 _____

審査委員 金 孝 根 _____

梁 熙 東 _____

申 璟 植 _____

梨 花 女 子 大 學 校 大 學 院

감사의 글

대학원에서의 2년은 많은 면에서 성장을 할 수 있었던 소중한 시간이었습니다. 그러한 성장은 제 부족함을 때로는 따끔한 충고로 때로는 따뜻한 격려로 함께했던 연구실 식구들과 교수님이 곁에 있어서 가능했던 일이었습니다.

바쁘신 가운데에도 이 논문이 완성될 수 있도록 끊임없는 지도로 바르게 이끌어 주신 신경식 교수님과 부족한 논문을 세심하게 검토해 주시고 심사해 주신 김효근 교수님, 양희동 교수님께 진심으로 감사 드립니다. 그리고 4학기 동안 정성껏 가르쳐 주시고 지도해 주신 경영학과 대학원의 모든 교수님들께도 감사 드립니다.

더불어 논문의 자료수집에 많은 도움을 주신 김명정씨와 승현언니, 논문의 시작부터 교정까지 늘 신경써주고 격려를 아끼지 않았던 용운오빠와 나의 친구 영은, 하연에게도 진심으로 감사의 말씀 전하고 싶습니다.

마지막으로 음으로 양으로 많은 힘이 되었던 우리 가족들에게 깊은 사랑과 고마운 마음을 전하며 이 작은 논문을 바칩니다.

목 차

논문개요.....	5
I. 서론	6
A. 연구의 목적	6
B. 연구범위 및 논문구성.....	7
II. 기업신용평가	9
A. 신용평가 개요.....	9
1. 신용평가의 개념	9
2. 주요 신용평가 모형	11
B. 소기업과 신용평가.....	15
1. 중, 소기업의 정의	15
2. 소기업의 역할	17
3. 소기업의 특성	17
4. 소기업 신용평가 모형	18
III. 연구방법	24
A. 통계분석 기법	24
1. 두 독립 정규집단의 비교: t-검정	24
2. 비모수적 검정	25
B. 의사결정나무(Decision Trees) 기법.....	29
1. 의사결정나무의 개념	29
2. 의사결정나무의 구성	29
3. 의사결정나무의 장, 단점	30
4. 의사결정나무분석의 알고리즘	30

IV. 실험설계	39
A. 데이터 수집방법 및 표본기업	39
B. 연구변수 및 연구방법의 설계.....	39
1. 연구변수	39
2. 연구방법의 설계	45
V. 실험결과	47
A. 입력변수의 선정- 단일변량 분석.....	47
1. 건전기업과 부실 기업의 평균 차이 비교 : t-검정	47
2. Mann-Whitney-Wilcoxon 검정	48
3. 카이제곱 적합도 검정	50
B. 의사결정나무에 의한 신용평가 모형.....	51
1. 의사결정나무의 구축 - Pruning전.....	51
2. 의사결정나무의 구축 - Pruning후.....	53
3. 건전/부도기업을 판별하는 Rule의 생성.....	54
4. 예측 정확도.....	56
VI. 결론	58
참고문헌.....	60
ABSTRACT	65

표목차

<표1> 신용평점모형 예시.....	13
<표2> 우리나라 중소기업의 범위.....	16
<표3> 체크리스트의 주요내용(예시)	20
<표4> 분할표.....	28
<표5> r X c 분할표(contingency table).....	31
<표6 > 분석대상 변수 및 세부내용.....	40
<표7> 의사결정나무에서 사용된 훈련용, 검증용 데이터.....	46
<표8> t-검정 결과 및 유의한 변수.....	47
<표9> Mann-Whitney-Wilcoxon 검정결과.....	49
<표10> χ^2 적합도 검정결과.....	50
<표11> Ruleset의 정리	54
<표12> 훈련용 및 검증용 데이터의 부도/건전 적중률.....	57

그림목차

<그림1> 소기업 경영평가를 위한 체크리스트(예시).....	20
<그림2 > 소기업 신용평가 모형구축 과정	45
<그림3> Pruning전 의사결정나무구조 및 가지별 데이터 수와 신뢰도	52
<그림4> Pruning후 의사결정나무구조 및 가지별 데이터 수와 신뢰도	53

논문개요

최근 금융시장의 경영환경이 급격하게 변화됨에 따라 금융기관 상호간의 경쟁이 치열해지고, 이전의 대기업 위주의 안정적인 여신운용이 어려워지게 되었다. 금융기관들은 종래의 대기업, 담보위주의 보수적인 대출제도에서 벗어나 정확한 신용평가에 의해서 신용 있는 차주를 개발하고 여신을 운용하여 경쟁력을 확보해야만 하는 입장에 처하게 된 것이다. 이에 일부 금융기관은 신용이 우수한 중, 소기업 및 가계에 대한 대출을 확대하여 위험을 분산시키고 수익성도 제고하는 전략으로 전환을 시도하고 있으며, 성공적인 전략 전환의 핵심이 되는 중,소 기업에 대한 정확한 신용평가의 중요성을 재인식하고 있다.

그러나 소기업의 경우 규모가 영세하고 담보능력에 한계가 있어 신용축적이 어렵고, 기업의 신용정보도 신뢰성과 투명성이 떨어져 신용평가를 위한 정보를 수집하기가 어렵다. 또한 일반 대기업이나 중견기업과는 달리 대출 수요가 많은데 비해 여신의 규모가 작아 상대적으로 심사비용이 높게 발생한다. 이 같은 소기업의 특성으로 인해 그동안 신용평가 모형을 구축하여 여신심사를 하는 경우가 드물었고, 대출 심사의 효과성과 효율성을 동시에 확보할 수 있는 소기업 신용평가 모형의 구축도 어려움을 겪고 있다.

이에 본 연구는 C은행의 소기업 고객들을 대상으로 고객 개요정보를 수집하였으며, 정성, 정량적 정보를 분석하는데 효과적인 의사결정나무 기법을 활용하여 소기업 신용평가 모형을 구축해 보았다. 더불어 문헌연구를 통해 신용평가와 소기업에 대한 전반적인 개념, 특징과 신용평가의 실태를 알아보았고, 모형의 구축에 활용될 통계기법과 의사결정나무기법에 대해 정리하였다.

그 결과 훈련용 데이터와 검증용 데이터 모두 73~79%의 안정된 예측 정확도를 보이는 소기업 신용평가 모형을 구축할 수 있었고, 이는 소기업의 특이성과 소기업 신용평가의 인프라 부족으로 모형 구축에 어려움을 겪고 있는 소기업 신용평가의 개선 가능성을 제시했다는 점에 의의가 있다.

따라서 이 같은 소기업 신용평가 모형의 구축과 적절한 활용은 앞으로 국내 은행의 다양한 시장에 대한 여신 마케팅 역량을 강화시켜 국내 금융기관의 경쟁력을 강화해 줄 뿐 아니라 궁극적으로는 국민 경제를 활성화하는데 기여할 수 있을 것이다.

I. 서론

A. 연구의 목적

신용평가는 기업의 채무 불이행 위험도를 측정하는 것으로, 은행 등 금융기관에 있어 대출여부의 결정 및 금리 등의 조건을 결정하는 중요한 근거가 된다. 그 동안 한국의 금융기관은 경제개발계획의 추진과정에서 여신의 만성적인 초과수요와 정부주도 하에 정책자금 위주의 인위적인 자금할당 덕분에 좋은 신용도와 높은 담보능력을 갖춘 대기업과 그 관계기업을 상대로 안정성 있는 대출에 치중할 수 있었다.

그러나 대기업, 담보위주의 대출관행은 IMF 경제위기, 대기업 부도, 금융권 자율화 등과 같은 급격한 환경 변화에 직면하게 되면서 심각한 부실 여신을 은행들에게 안겨 주었고, 이 여신위험을 효과적으로 감소시키지 못한 금융기관은 결국 인수 및 합병을 통한 구조조정을 하기에 이르렀다.

최근의 이러한 금융환경변화는 금융기관들로 하여금 정확한 신용평가를 통한 대출의 중요성을 깊이 인식시키는 계기가 되고 있으며, 대기업의 탈은행화¹ 경향 또한 안정적인 담보력을 갖춘 기업을 대상으로 한 여신운용이 더 이상 용이하지 않음을 의미한다. 따라서 일부 금융기관들은 우수한 중, 소기업 및 가계에 대한 대출 확대의 중요성을 인식하고 이를 통해 위험을 분산시키고 수익성도 제고하는 방향으로 대출 전략을 전환하고 있다.

그런데 소기업의 경우에는 신용정보의 신뢰성이 매우 떨어져 재무정보를 바탕으로 한 객관적인 신용평가가 어렵고 대출 수요가 많은데 비해 여신의 규모가 작아 심사비용이 많이 든다는 특징이 있다. 때문에 대출심사를 합리화 하기 위해서는 이에 대한 충분한 연구를 통해 소기업 신용평가모형을 구축하여 심사의 프로세스의 효과성과 효율성을 높일 필요가 있다.

그 동안의 신용평가 모형 구축에 대한 연구를 살펴보면 상장기업 또는 대기업을 대

¹ 2,000년 말까지 대기업의 부채비율이 200% 이하로 낮추기 위해 대기업들은 자산의 매각, 차입금 및 회사채의 상환, 증자 등의 활동을 하고 있고 외환 및 자본시장 자유화가 확대되면서 일정한 신인도를 갖춘 대기업은 국제 금융시장에서 직, 간접 금융수단을 활용할 것으로 국내금융시장 의존도는 하락하게 되는 현상을 대기업의 ‘탈은행화(脫銀行化)’ 현상이라 한다.

상으로 재무비율을 이용한 통계적 부도예측모형이나 인공지능기법(AI기법)을 이용한 부도예측모형에 대한 연구(Beaver,1966; Altman, 1968, 1983; Gentry et al., 1985;; Edminster, 1972; Elam, 1975; Kaplan and Uriwitz, 1979, Ohlson, 1980; Zmijeswski, 1984, 이견창 등, 1994; Barniv et al.,1997; Bell, 1997; boritz and Krnnedy, 1995; Jo et al., 1997)가 주류를 이루고 있다. 또한 중소기업을 대상으로 한 신용평가 모델 구축에 관한 연구라 하더라도 대부분 비교적 규모가 큰 중견기업을 대상으로 한 연구여서 (이재식,한재홍, 1995; 오갑수,1996; 김철교, 백용호, 1997; 이장배, 1998; 고영배, 1997; 이훈영, 조옥래, 이시환,1998), 이 모형을 소기업에 적용하는 데는 무리가 있다.

소기업의 신용평가 시 중요한 비재무 항목을 반영한 신용평가에 대한 연구 역시 미진하여, 중기업 신용평가에 있어서 비재무 정보의 유용성을 검증한 연구(김충섭, 남기전, 1997; 김충환, 1997; 노상월, 1997; 정연갑, 1997; 신우규, 1998; 이진우, 2000; 정인수, 1999)정도가 있을 뿐이다.

이 같은 상황은 소기업의 신용평가를 실제로 수행하고 있는 금융기관도 다르지 않아, 소기업 신용평가에 대해 전문적 경험과 고객 데이터 등의 인프라 부족으로 소기업 신용평가 모형 구축에 어려움을 겪고 있다.

따라서 본 연구는 C은행의 소기업 고객들에 대한 질적정보를 survey를 통해 직접 수집하였고, 이를 바탕으로 정량, 정성정보를 처리하는 데 효과적인 의사결정나무 기법을 활용한 소기업 신용평가 모형을 구축하고자 한다.

B. 연구범위 및 논문구성

본 연구는 의사결정나무(Decision tree)를 이용하여 소기업 신용평가 모형을 개발하는데 그 목적이 있다. 이를 위해서 국내외 문헌 연구를 통해 그 동안의 신용평가제도 및 신용평가 기법에 대한 연구와 비재무 정보를 이용하여 신용평가에 적용한 연구들이 어떻게 이루어 졌는지, 그리고 현재 우리나라 소기업 신용평가가 어떻게 이루어 지고 있는지에 대해서 살펴보았다. 또한 본 연구에서 사용될 의사결정나무 기법, 비재무 항목의 선정에 사용된 통계기법들에 대해 연구, 조사한다.

본 연구에서는 소기업 개요정보를 기반으로 한 기업의 비재무정보에서 단일변량 분석 통계(t-검정, Mann-Whitney U 검정) 을 이용하여 유의한 항목을, χ^2 기법을 이용한 변수들의 적합성을 검증하여 유의성과 적합성을 만족하는 변수들을 선정한다. 이후 선정된 변수들을 의사결정나무의 입력변수로 사용하여 모형을 세우고, 이를 소기업의

건전/부실을 판별하는 소기업 신용평가 모형으로 활용하고자 한다.

제 I 장의 서론에서는 연구의 목적과 연구의 범위 및 구성에 대해 기술한다.

제 II 장에서는 기업신용평가에 대한 전반적인 개념과 소기업과 기존의 소기업 신용평가모델에 대해 기술한다

제 III 장 연구방법은 실험에서 활용 될 통계기법 및 의사결정나무기법에 대해 구체적으로 기술한다.

제 IV 장 실험설계에서는 실험에 사용된 데이터 수집방법과 표본기업을 정의하고, 통계적 분석 기법을 이용하여 입력 변수를 선정한 후, 의사결정나무 기법에 의한 소기업 신용평가 모형을 구축한다.

제 V 장 실험결과에서는 변수들의 통계적 분석 결과와 이 변수들을 입력하여 얻은 의사결정나무 모형의 실험결과를 제시하고, 모형 및 검증용 데이터의 적중률을 비교한다.

제 VI 장 결론에서는 연구의 성과 및 한계점을 기술하고 향후 연구과제에 대해 고찰한다.

II. 기업신용평가

A. 신용평가 개요

1. 신용평가의 개념

가. 신용평가의 의의 및 목적

신용이란 믿음 또는 신뢰성을 의미하나 이를 경제거래와 관련지어 본다면 자금거래 또는 실물거래에서 기업이 금전이나 재화를 정해진 기간 내에 당초의 약속대로 상환, 지불 또는 변제할 수 있는 능력이라고 정의할 수 있다. (손정식, 1998) 즉, '신용이란 자금을 차입하는 자가 그 차입한 자금을 제 기간 내에 갚을 수 있는 능력'이라 할 수 있다.

일반적으로 신용평가는 금융기관 대출의 회수불능위험을 평가하는 것이다. 금융기관의 기업신용평가는 금융기관여신운용 세칙, 제2조 여신심사에서 '금융기관은 여신 취급 시 차주의 재무상태, 타 금융기관으로부터의 차입금, 기타 채무를 포함한 전반적인 금융기관상황, 사업현황 및 전망과 여신의 용도, 소요기간 및 상환능력 등의 정확한 분석을 통하여 적정을 기하여야 한다'라고 규정하고 있다.

신용평가는 기업내외의 경제사정, 즉 재무상태, 경영능력 등의 내부상황과 동일업종의 경기전망, 정부방침 등 외적여건을 조사하고 이를 바탕으로 기업의 신용도(credit worthiness)를 결정하는 것으로 은행에서는 대부분 여신거래처의 신용을 조사하는 것을 원칙으로 하고 있다.

즉, 기업신용평가는 신용의 주체인 채권자와 채무자 사이의 장래 신용관계를 파악하기 위하여 신용조사 과정을 통해 수집된 기업의 경영활동상태나 결과 및 계획 등을 정리, 분석, 비교, 검토하여 그 기업의 현재 신용도를 평가하고 신용정보 수요자에게 제공하는 것을 말한다.

신용평가 결과는 금융기관에 대한 여신정보의 제공으로 효율적인 여신업무수행에 도움이 된다. 기업에 대해서는 자본조달기회의 제공으로 재무구조 개선에 기여하게 되고 투자자에 대해서는 투자에 대한 위험정보를 제공하게 된다. 또한 이러한 정보는 신용평가 정보이용자가 보다 기업의 상황을 이해하고 합리적인 의사결정을 하는데 도움이

되도록 함으로써 금융시장의 건전한 육성 및 자금배분의 효율성을 제공하는데 기여하며, 나아가 신용경제사회를 구현하는데 그 목적이 있다.

기업의 신용도에 따라 신용도가 좋은 기업은 보다 저렴하고 질이 좋은 자금을 용이하게 조달할 수 있다. 반면, 신용도가 불량한 기업은 자금조달에 제한을 받고 위험프리미엄의 지불로 자금비용이 상승하게 된다. 따라서 신용평가제도는 금융자금의 수요와 공급을 효율적으로 결합시키는 역할을 수행하게 된다.

또한 기업의 신용상태에 따라 기업의 자금조달능력이 결정되게 되므로 기업이 자신의 신용관리를 철저히 하도록 하여 기업의 재무구조 개선을 유도하며, 정보력이 떨어지고 분석능력이 약한 투자자들에게 사전에 투자위험의 정도를 알려주어 투자자를 보호한다. 이는 일반투자가 및 기관투자자의 자본시장 참여기회를 확대시켜 자본시장의 발전을 촉진하는 역할을 하게 한다.

나. 신용평가의 기능과 역할

신용평가결과는 다음과 같은 세가지 부분에서 그 기능과 역할을 수행한다.

(1) 투자자 보호

신용평가의 주된 역할은 투자자들이 이미 보유하고 있거나 매입하고 있거나 매입하려고 하는 채권의 잠재적인 채무불이행위험에 대한 비교가능하고 일관성 있는 척도를 계속적으로 제공하여 합리적인 투자결정에 기여한다. 또한 유가증권 중개기관에 대해 유가증권 인수 할인에 대한 위험을 감소시켜 유가증권의 매매를 촉진시키는 등 자본시장의 발전에 기여한다.

이러한 신용평가는 투자자들에게 투자손실을 보상해 주지는 못하지만 그러한 투자손실 가능성을 스스로 관리할 수 있도록 도와주는 기능을 수행한다.

(2) 금융기관의 여신운용 참고지표

신용평가는 은행 등 금융기관의 우량기업체 발굴, 여신 지원여부 및 지원조건 등에 참고지표로 활용되며, 기업의 신용도에 따른 금리의 차등 운용으로 금융자율화 및 자금의 효율적 배분에 기여한다. 신용의 창출자, 전달자 및 감시자로서 신용정보 확충을 통하여 신용을 중시하는 풍토를 조성하고 신용경제사회 정착에 이바지한다.

(3) 기업의 자금조달 원활화

신용평가는 기업에 대하여 자기신용에 적합한 자금조달 기회를 제공하며, 신용등급은 유가증권의 발행조건을 적절하게 결정하는 지표로서 자금조달비용 등 참고자료로 활용한다. 신용평가가 금융시장 전반에 걸쳐 활성화되면 시장에서 신용위험과 관련한 정보의 유통을 원활하게 하여 투자자에게 보다 다양하고 많은 정보를 제공하면서도 자율적이고 건전한 자본시장의 발달을 촉진하게 하여 자금조달 가능성을 증대 시키고, 조달비용을 절감시키는 효과가 있다.(김형수, 1999)

2. 주요 신용평가 모형

신용평가를 위한 기업 부실화 예측모형으로는 단일변량 분석과 다변량 분석 모형이 있으며, 다변량 분석은 다변량 판별분석, 다중회귀분석, Logit 및 Probit 분석 등이 있다. 최근에는 요인분석, 이상치 방법(주관적방법), 인공지능기법, 의사결정나무모형 등이 거론되고 있으나, 본 연구에서는 실무적으로 이용되고 있는 계량모형과 신용평정모형 및 전문가 판단모형 위주로 고찰하려고 한다.

가. 계량모형

(1) Z모형

다변량 판별분석에 의한 기업실패예측에 관한 Altman의 모형(1968년)으로서 평가대상 기업들을 건전, 부도기업 그룹으로 분류한 후 이들 그룹을 차별화 할 수 있는 판별계수를 구하였다. 판별함수를 이용하여 개별기업의 신용상태를 점수화 한 후 각 기업의 신용평점을 두 그룹의 평균과 비교하는 기법으로서 최근까지 기업 도산 예측이나 평점제도의 근간으로 사용되고 있는 유용한 기법이다.

Altman은 표본을 추출하여 22개의 재무비율을 분석대상으로 하여 여러 가지 판별함수(discriminant function)조합을 만들었다. 판별분석의 효율이 하락되지 않는 판별함수를 제거하고 최종 판별함수에 포함될 변수를 5개의 재무비율(기업의 유동성, 수익성, 레버리지, 지급능력, 활동성)로 축소하였다.

$$Z = 0.012X1 + 0.014X2 + 0.033X3 + 0.006X4 + 0.999X5$$

X1: 운전자본/ 총자산(%)

X2: 유보이익/ 총자산(%)

X3: 이자 세금차감전 이익/ 총자산(%)

X4: 자본의 시장가치/ 총부채의 장부가격(%)

X5: 매출액/ 총자산(회)

Z: 판별값

이를 이용하여 판별 함수식으로 표현되는 기업부도예측모형(Z-score모형)을 위와 같이 만들었으며, 판별함수에 의한 판별점을 기준으로 실패기업, 건전기업, 무지의 영역으로 구분하는 방법이다.(Edward I. Altman, 1968)

(2) ZETA 모형

Altman은 1977년 Harldeman과 Narayanan등과 더불어 Z-score모형을 개선하여 ZETA모형을 개발하였다. 이 모형은 27개 주요 재무변수 중 7가지 변수를 독립변수로 선정(투자수익율, 이익의 안정성, 지급능력, 누적적 수익성, 유동성, 자본구조, 기업규모)하고 다변량 판별분석을 실시하여 기업 실패예측모형을 나타내는 것이다. 이는 기업의 신용상태에 관한 사전적 정보와 잘못된 판단에 따른 코스트(1종 오류: 부도기업을 건전기업으로 판단하는 오류, 2종 오류: 건전기업을 부도기업으로 판단하는 오류)까지 고려하여 Z점수모형을 보다 개량, 일반화 시킨 것이다. (Edward. I. Atlman, Robert G. Heldeman, P. Narayanan, 1977)

(3) Logit 모형

Logit 모형은 Ohlson의 기업부도예측모형에 관한 초기의 연구로서 기존의 연구들이 사용했던 다변량 판별분석의 통계적 문제점과 표본선정에 있어서의 문제점을 지적하고, 이를 해결하기 위해서 부실예측을 통계적 방법으로 해결한 모형이다. Logit 분석 기법은 전체 평가대상 기업의 신용위험 확률을 추정한 후 추정치를 이용하여 개별기업의 신용위험 확률을 계산하고 이것이 일정수준을 넘으면(절대평가 기준을 활용) 부도기업으로 분류한다. Logit모형은 기존의 다변량 판별분석과 중회귀분석과는 달리 각 판별 집단의 예측에 대한 분산-공분산 행렬이 동일하고 예측치의 분포가 정규분포를 갖는다는 가정을 충족시켜야만 하는 통계적 한계점을 해소할 수 있다.

Logit 모형은 종속변수에 0 또는 1이라는 질적 변수를 두고 설명변수에는 양적인 변수들을 두어(기업이 부도가 나면 1, 부도가 나지 않으면 0) 분석하고 있으며, 종속변수가 부도확률을 나타낼 수 있다. 평가모형을 개발하는 연구자나 추출된 자료를 분석하는 분석가, 보고내용을 가지고 의사결정을 해야 하는 경영진 모두에게 매우 유용한 도구가 될 수 있다. (J.Ohlson, 1980)

(4) 인공신경망 모형

평가대상 기업들의 신용정보를 입력시키면 반복적 학습장치를 통하여 기업의 신용상태를 인식하는 기법이다. 기업 신용평가의 정확성 및 객관성을 높이기 위하여 과학적 평가모형 및 기법을 적용하여 왔으며, 최근 컴퓨터의 발달과 함께 1980년대 후반부터 기업신용평가, 도산예측 및 주가예측 등에 응용하는 연구들이 진행되었다. 이 연구들에서는 기존 통계 방법론과의 비교 연구를 통해 인공지능기법의 우수성을 증명하고 있다. (Elmer and Borowski, 1998; Odom and Sharda, 1990; Cadden, 1991; Raghupathi et al., 1991; Tam and Kiang, 1992; 신경식과 한인구, 1998; Lee, 1993; Jo et al., 1995) 또한 최근에는 예측력을 향상시키기 위한 인공신경망 통합방법론을 제시하는 연구들도 발표되었다.(신경식과 한인구, 1998; 신경식 등, 1998)

나. 신용평점 모형

신용평점모형은 재무항목과 비재무항목들의 상대적 중요도에 따라 배점을 부과한 후 평가된 점수를 합계하여 총 평점을 구하는 방식이다. 업종별 재무비율분석을 통하여 해당기업의 재무구조 건전성(안정성, 수익성, 활동성, 생산성, 성장성)을 분석하고 비재무항목을 통하여 사업성, 거래신뢰도, 경영능력 등을 종합적으로 평가한다. 중소기업의 경우 소기업일수록 비재무항목의 배점을 높이고 중기업의 경우 재무항목과 비재무항목의 배점을 비슷하게 두는 것이 일반적이다.

<표1> 신용평점모형 예시

구분	평가항목	배점	A	B	C	D	E
안정성	자기자본비율						
	유동비율						
	차입금의존도						
수익성	매출액경상이익율						
	총자본순이익율						
활동성	영업자산회전율						
현금흐름	이자지급후 CF/부채						

	외부차입전 CF/부채						
성장성	매출액증가율						
생산성	부가가치율						
재무평점							
비재무평점							
종합평점							

※ 윤태순, 1999.8, "국내은행의 신용평가 모형 개선과 효과적인 Credit Rating System 구축방안", 월간조흥경제

재무항목의 경우 모든 업종 및 규모에 같이 적용할 것이 아니라 업종, 규모별로 유의성이 있는 재무항목을 선택하여 배점을 차별화 할 필요가 있다. 재무자료의 분석회계 가능성과 신뢰성을 확인하기 위하여 자료의 검정 및 수정작업이 이루어져야 할 것이다. 중소기업의 생산, 영업활동 및 재무적 특성을 감안하여 기업을 업종별, 그룹별로 분류하고 신용평가 시 건전, 부실기업 판별의 통계적 유의성 및 예측력, 관련변수들의 상관관계 등을 고려하여야 한다. 최종적으로는 선택된 변수들의 통계적 분포를 용도에 맞게 등급별 구간을 설정하여야 한다. 또한 다변량 판별분석 등을 통하여 재무변수들 간에 상대적 중요도를 결정하고 이에 해당하는 배점을 할당한다. 이와 같은 과정을 통하여 업종별로 설정된 모형은 시범운행을 통하여 문제점을 보완하고 적합성을 높이는 피드백작업을 통하여 개선된다. (손상호, 김동환, 1999)

우리나라의 경우 한국은행의 재할인율을 받은 은행은 한국은행에서 정한 기업체 종합평가 작성기준에 의하여 평점을 산출하여 거래 적격성을 판정한다. 항목별 평가등급 기준은 매년 개정하여 변동되는 경제여건을 감안하고 있다. 이 평가기준표에 의한 평점이 50점 이상이면 적격업체이며, 80점 이상인 경우에는 건전기업으로 판정하고 있다.

다. 전문가 판단모형

전문가 판단모형은 신용평가에 전문지식을 가진 평가자가 계량화하기 어려운 기업정보를 획득하여 기업의 실체를 파악하고 논리적 추론 및 설득력 있는 결론을 도출하는 평가모형이다. 소기업에 대한 비재무적 정보는 수치화, 계량화가 어려운 정보로서 평가자들의 주관적 판단을 객관적 논리 및 자료로 평가해야 한다.

비재무항목은 영업전망, 시장성, 경쟁력, 금융거래신뢰도, 경영능력, 재무적 융통성 등 주관적 판단사항이 대부분이다. 이를 점수화 할 경우에는 각 항목의 중요성을 경험적으로 파악함은 물론 객관적인 논리 및 자료를 통해 항목간의 관계(인과관계, 병렬관계, 우선순위 등)를 분석하고 재무항목에 할당된 점수와의 상대적 관계를 고려하여 가중치(배점)를 조정하여야 한다. 중,소기업은 대부분 회사의 자산 부채를 재무제표에 제대로 반영하지 않거나, 자료의 신빙성이 확보되지 않으므로 기업의 행태, 재무상황 등에 대한 전문가의 판단이 중요하다.

전문가 판단모형의 일종인 보증기금 표준심사기준표 (3억원 초과 보증거래기업 보증심사에 활용)의 내용은 대부분 주관적 사항의 기술로 이루어지고 있으며 그 예시를 보면 다음과 같다.

(신용보증기금 표준심사기준표 내용)

- ① 경영능력 검토: 품성 및 건강, 기업관, 지도력, 예견력 및 위기관리, 동업계 중사경험 및 능력, 평판
- ② 경쟁력 검토: 업력 및 사업장 현황, 유망성, 원재료 조달, 판매상황, 품질 및 기술수준, 인력개발수준
- ③ 금융거래상황 검토: 차입금 규모 및 추이, 거래신뢰도
- ④ 재무상태 검토: 매출액 및 지급어음 규모, 상환채무 현황
- ⑤ 소요운전자금 검토: 현금흐름표 검토
- ⑥ 부실징후 검토: 운전자금 회전기간, 차입금규모, 재고기간, 자금여력도, 현금흐름, 활동밸런스를 평점화하여 표시

신용보증기금의 심사기준표는 주관적 사항이 매우 강하게 반영된 종합판단시스템으로서 기업의 신용도가 우량하거나 미흡하여도 심사자 및 전결권자의 책임 하에 상기 심사내용을 주관적으로 판단하여 보증지원 가부 및 지원규모를 결정하기 때문에 객관적인 기준보다는 주관적인 판단이 강조된 신용평가제도이다. (이윤재, 1997; 손상호, 김동환, 1999)

B. 소기업과 신용평가

1. 중, 소기업의 정의

중, 소기업의 개념을 정의할 때 그 기준은 나라별, 업종별 상황에 따라 각기 다르다. 일반적으로 중, 소기업(small and medium size business)이란 대기업과 비교하여 상대적으로 작은 규모의 기업을 말하는 것으로, 기업규모나 매출액 규모로서 기업을 분류할 때도 상대적인 의미를 갖는 것이다. (조희영, 박상범, 1997) 그러나 중, 소기업을 정의할 때는 종업원수, 총자산, 자본금, 매출액 등의 양적 기준과 경제 지원정책, 기업의 사업활동형태, 생산품의 라이프 사이클, 기업경영의 동기 등과 같은 질적 기준이 이용되지만 중소기업의 정의를 단적으로 구분하기는 매우 어렵고 대개의 경우 질적 특성을 무시하지 않는 범위 내에서 양적 특성을 고려해야 한다는 것이 일반적이다. (박상범, 1998) 중소기업에는 이질적이고 다원적인 역사적 제 요소가 혼재되어 있어 중소기업의 개념을 획일적으로 규정하기 보다 각국의 경제적, 정치적 사정에 따라 적절하게 규정하고 있다.

그런데 실정법상 규정된 중, 소기업은 종업원수, 총자산, 자본금, 자기자본 및 매출액 등 양적인 여러 기준에 의하여 그 범위가 정해지고 대기업과 중소기업간의 규모가 구분되고 있다. 이러한 양적 기준은 각국의 기준에 따라 다르게 정의되고 있으며, 일본에서는 중소기업을 종업원수와 자본금을 기준으로 정의 하고 있고, 미국의 경우 종업원수와 매출액을 중소기업정의의 기준으로 사용하고 있다. (Clifford. M. Baumbach, 1985) 우리 나라의 경우는 중소기업기본법에서 일본과 마찬가지로 종업원수와 자산규모를 양적 척도로 하여 중소기업의 규모를 설정하고 있다. (한희영, 1985)

우리 나라에서는 양적 및 질적 척도에 의해 중소기업을 정의하고 있는데, 최초로 중소기업에 대한 개념이 규정된 것은 1961년 7월에 제정된 중소기업은행법이다. 그 당시의 법에 의하면 중소기업이란 상시종업원이 5인 이상 100인 이하, 광업은 200인 이하이며, 총자산액이 2천만원 이하의 규모로서 제조업 또는 광업을 경영하는 자영인 또는 법인으로 규정하였다.

그러나 1961년 제정 공포된 중소기업협동조합법과 1966년에 중소기업기본법이 제정, 공포되는 과정에서 중소기업에 관한 규정이 완화되고 중소기업 관계법이 수차례 개정되어 왔다. 현행우리나라 중소기업의 정의는 1995년 7월에 개정된 중소기업기본법 제2조에 근거한 내용을 토대로 하고 있는데, 그 내용은 아래의 <표2>와 같다.

<표2> 우리나라 중소기업의 범위

구분	1982	1995년 이전	개정 후
----	------	----------	------

		소기업	중기업	소기업	중기업
제조업 광업 운송업	상시종업원 300인 이하 또는 자산총액 5억원 이하	상시종업원 20인 이하	상시종업원 21~300인 이하	상시종업원 50인 이하	상시종업원 51~300인 이하
건설업	상시종업원 50인 이하 또는 자산 총액 5억원 이하	상시종업원 20인 이하	상시종업원 21~200인 이하	상시종업원 30인 이하	상시종업원 31~200인 이하
상업 및 서비스업	상시종업원 20인 이하 또는 자산 총액 5천만원 이하	상시종업원 5인 이하	상시종업원 6~20인 이하	상시종업원 10인 이하	상시종업원 11~20인 이하

※출처:중소기업 기본법: 1995.1.5 법률 제 4897호- 중소기업기본법 시행령 제 2조
개정: 1995.7.1

2. 소기업의 역할

단일 소기업은 대기업이나, 중기업에 비해 자본, 매출액, 부가가치, 종업원 수와 같은 규모면에서는 작다고 할 수 있으나, 중, 소기업의 전체적인 측면에서 본다면 대기업에 비하여 고용수준과 생산수준, 부가가치 측면에서의 그 비중을 무시할 수 없다. 특히, 소기업은 도소매, 서비스업 등 비 제조업 분야에서 사업체수와 고용의 많은 비중을 차지하고 있다.

또한 이들이 성장하면 결국 중기업이나 대기업이 될 수 있고 중기업이나, 대기업의 모체가 될 수 있다. 현대와 같이 다양화를 강조하는 시대에서는 환경에 대한 신속하고 유연한 대응력과 창의력을 갖추고 있는 소기업과 같은 기업형태가 더욱 효과적일 것이다. 따라서 소기업의 이 같은 특성들을 더욱 발전시킨다면 향후 국민경제에 소기업이 새로운 역할을 수행할 가능성이 무한하다 할 수 있다.

3. 소기업의 특성

첫째, 자본규모의 영세성으로 사업규모 및 범위가 제한적이며, 금융기관으로부터 담보나 신용의 부족으로 여신서비스를 받는데 어려움이 있다. 따라서 대자본에의 종속성이 강하여 많은 기업이 모기업과 하청계열체제를 통하여 종속관계를 이루고 있다. 이

러한 대자본에의 종속관계는 대자본과 중소기업간에 대등한 거래관계를 왜곡시키는 현상을 낳게 한다.(박상범, 1998)

둘째, 사업이 대개 가족을 중심으로 운영되는 경우가 많아 기업의 지속성이 결여될 수 있으나 조직이 간소하여 조직의 비대함으로 인해 생길 수 있는 구조적 모순이나, 비능률을 쉽게 제거할 수 있는 반면 유연성과 신속성을 갖을 수 있기도 하다.

셋째, 회계 재무제표나 서류체계가 대기업에 비해 미흡하고 무계획적인 자금관리로 재무정보에 대한 신용도가 낮아 자본 동원에 한계가 있어 주로 개인적인 신용에 의하여 자본을 조달하여야 한다. 이 같은 개인신용에 의한 자본조달은 대규모 자본을 형성하는 데 큰 어려움이 되고 있으므로 정부차원의 중, 소기업 육성을 위한 장기적인 자금 지원정책이 요구된다.

4. 소기업 신용평가 모형

가. 소기업 신용평가 모형의 구축 시 고려해야 할 특징

소기업은 신용평가 시에 재무제표에 대한 신뢰성이 떨어져 재무정보 보다는 비재무항목을 적극 발굴하여 활용해야 한다. 특히 비재무항목에 대한 신용평가는 평가자의 주관적 판단에 의해 좌우될 뿐만 아니라 판단된 주관적 요인을 객관적인 데이터로 전환하기가 용이하지 않아서 계량모형과 평점모형에 충분히 반영하기가 힘들다.

소기업 여신은 기업여신과 소비자 여신의 특성이 혼재되어 있고, 재무정보에 대한 낮은 신뢰도와 대표자 개인의 능력과 특성에 의해 신용도가 좌우되는 특성을 지니고 있으며, 다수의 업체를 대상으로 소액의 여신을 취급해야 하기 때문에 심사비용을 최소화할 수 있도록 효율성측면을 반영해야 한다. 또한 소기업은 일반적으로 외부 환경 여건에 의해 크게 좌우 되기 때문에 평가시점의 경기상황에 따라 과대 또는 과소 평가되는 경향이 있다는 점도 신용평가 시 고려해야 할 것이다.

나. 기존 소기업 신용평가 모형

(1) 전문가 판단모형을 이용한 소기업 신용평가

소기업은 대부분 재무제표를 구비하고 있지 않거나 동 자료의 신뢰성이 확보되지 않으므로 기업의 행태 및 재무상황 등에 관한 전문가의 판단이 중요하다. 따라서 소기업의 신용 평가 시 사용되는 재무관련 데이터는 매출액과 각종 차입금 정도 등으로 대별

될 수 있으며, 비재무항목은 대표자의 경영능력과 평판, 주요거래처 상황 등에 의해 특징지어진다.

차입금의 경우 기업이 제출한 금융거래 확인서를 통해 당좌대출, 운전자금차입금, 장기차입금, 사채의존도 등을 확인한다. 대표자의 경영능력과 평판의 경우 경영자의 경영전념도 및 자질, 경험 정도, 경영자에 대한 세평(거래업체 및 금융기관의 평판), 경영안정성 등의 항목을 종합적으로 평가할 필요가 있다. 주요 거래상황은 주요 고정거래처의 확보, 주거래처의 신용상태, 판매처 다변화, 제품에 대한 평판 등을 종합적으로 평가할 필요가 있다. 특히 중소기업의 경우 불량거래자에 대한 판단이 중요하므로 각종 공과금(전화료, 전기료, 수도료 등), 이자, 임금, 신용카드대금 등의 연체, 전과기록, 국세체납정보, 최근 당좌부도 여부, 신용불량정보, 경영주 소유부동산에 대한 압류, 가처분 등의 확인이 필요하다. 또한 부실징후에 대한 판단을 위해 최근 차입금 의존도의 증가 추이, 매출액 감소 추이, 총차입금(또는 운전자금차입금)이 매출액(또는 자기자본)에서 차지하는 비중 추이, 사채의존도 추이 등의 점검이 필요하다.

(2) 전문가 판단모형의 예- 체크리스트를 통한 평가

소기업의 평가 시 기존의 평점모형을 활용하는 방법과 전문가 판단모형의 간단한 예로서 단계적으로 체크리스트를 통해 평가하는 방법을 고려할 수 있다. 심사자에게 책임과 권한을 최대한 부여할 경우 체크리스트를 통한 평가가 적극 활용되고 있다.

그러나 체크리스트를 활용한 신용평가는 반드시 다음과 같은 사항들에 대한 고려가 있어야 할 것이다.

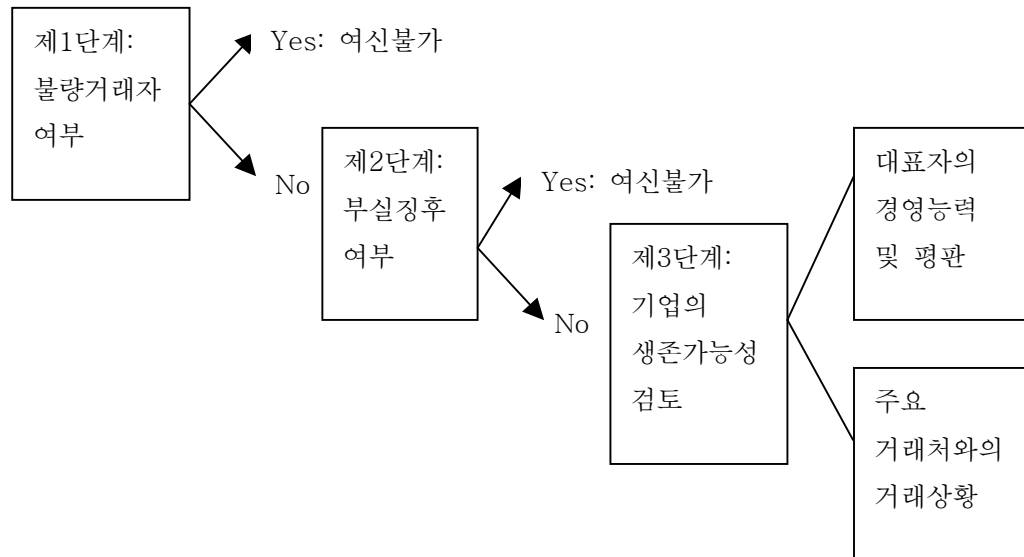
첫째, 객관적 데이터가 취약한 상황에서 평점모형은 항목 추출 및 배점간의 자의성으로 인해 총평점의 신뢰성에 문제가 발생할 수 있는 한편 평가자는 평점에 의존하여 책임을 회피하려는 도덕적 해이문제를 야기할 수 있다.

둘째, 체크리스트를 통한 평가는 체크리스트를 통해 평점을 계산하는 것이 아니라 평가자가 해당항목의 특기사항과 중요사항만을 기술한 뒤 최종적인 판단은 이들을 감안한 평가자의 주관적 판단에 의존하는 방식이다.

셋째, 이와 같은 방식의 부작용을 최소화하기 위해서는 평가자의 전문성이 전제가 되어야 하고 이들에게 최대한의 권한과 책임을 부여하는 동시에 사후적인 평가를 통해 보상할 수 있는 시스템이 갖추어져야 한다.

아래 <그림1>와 <표3>는 한국금융연구원에서 제시하고 있는 소기업 경영평가를 위한 체크리스트의 예시와 체크리스트의 주요내용에 대한 예시를 각각 보여주고 있다.

<그림1> 소기업 경영평가를 위한 체크리스트(예시)



※출처: 손상호, 김동환, 1999, “중소기업 신용평가제도 개선방안”, 한국금융연구원

소기업은 관련 데이터와 신용이 아직 축적되어 있지 않으므로 불량거래 및 부실징후 여부의 판단부터 시작하여 경영자와 거래처에 대해 평가하는 것이 바람직하다.

우선, 불량거래자 여부부터 판단하게 되는데, 일단 불량거래가 범주에 포함될 경우 그 다음 단계의 판단이 유보되면서 여신대상에서 제외된다.

불량거래자가 아닌 경우, 부실징후 여부를 판단하여 부실징후의 범주에 포함되지 않으면 그 다음 단계에서 해당기업의 생존가능성을 검토하게 된다.

우선 기업의 업력이 짧고 규모가 작으면 대부분 소유, 경영자 1인에 의해 좌우되는 경우가 많으므로 이들의 능력과 경영방침 및 평판을 평가하는 것이 중요하다. 또한 소기업은 관련 거래 기업들로부터 많은 영향을 받으므로 주요 거래처와의 거래관계 및 신용유지 상태, 그리고 제품에 대한 평판이 중요하다.

<표3> 체크리스트의 주요내용(예시)

불량거래자 여부	<ul style="list-style-type: none"> • 최근 연체대출금 보유여부, 당좌부도 여부, 신용불량 정보보유 • 최근 주요거래처와의 거래중단 여부
-------------	---

부실징후 여부	<ul style="list-style-type: none"> • 총차입금이 당기 매출액 초과하는지 여부, 운전자금 차입금액이 당기 매출액 50%초과 여부 • 최근 차입금 의존도 및 증가세 추이, 매출액 감소추이 • 사채의존도 추이, 최근 3년간 적자여부, 납입자본금 잠식여부, 부채비율 수준
대표자의 경영능력 및 평판	<ul style="list-style-type: none"> • 최근 대표자 및 핵심 임직원 변경여부, 경영자 세평(거래업체 및 금융기관의 평판) • 경영자의 경영전념도 및 지식, 경험 정도(동업계 종사기간) • 경영안정성, 자가사업장 보유 여부
주요거래처 와의 거래상황	<ul style="list-style-type: none"> • 주요 고정거래처 확보, 주요 거래처와의 신용상태 • 판매처 다변화 및 제품에 대한 평판

※출처: 손상호, 김동환, 1999, “중소기업 신용평가제도 개선방안”, 한국금융연구원

다. 소기업 신용평가 시 비재무정보의 유용성

(1) 재무정보에 의한 소기업 신용평가의 한계

우리나라 기업 회계기준에서는 회계정보 이용자가 기업실태와 관련하여 합리적인 의사결정을 할 수 있도록 재무상의 자료를 일반적으로 인정된 회계원칙에 따라 처리하여 유용하고 적정한 정보를 제공하는 것을 목적으로 한다라고 규정하고 있다. 이러한 목적을 가진 회계정보가 지금까지 은행 및 신용평가 기관의 신용평가나 기업 부실에 있어서 여타 정보보다 많이 이용되어 왔던 것은 회계정보의 객관성과 정보획득의 상대적 용이성 등을 고려한 결과이다. 그러나, 신용평가나 부실 예측 시 지나치게 회계정보에만 의존하는 것에 대해서는 비판적 견해가 많다. 특히 “주식회사의 외부감사에 관한 법률”에 의해 공인회계사의 결산 감사를 받는 외부감사대상 기업보다는 중소기업의 경우 더 심각한 것으로 인식되고 있다. 이들 비판을 요약하면 다음과 같다.(김용기,1983)

첫째, 재무제표상의 공표이익이 절대적인 것은 아니라는 점이다. 회계처리의 자의성 및 임의성에 따라 서로 다른 회계처리에 의하면 이익은 제각기 달라질 수 있기 때문에 공표이익은 경영정책 등을 감안하여 선택된 회계처리에 의하여 산출된 상대적인 것이 지 절대적인 것이 아니다. 더욱이 외부감사를 받지 않는 중소기업은 그 자의성이 더욱 확대된다.

둘째, 재무제표에 계상되지 않는 기업정보가 있다. 비계량적인 기업정보 영업권, 상품권, 기술수준, 타기업과의 관계, 동업계의 지위, 시장성과 마케팅 능력 등 상당한 중요정보가 재무제표에 표시되지 않는다.

셋째, 재무제표가 의식적, 무의식적으로 분식되고 있다. 특히 기업의 의도적인 분식 결산에 따라 재무제표가 분식 되었을 경우에는 그 분식 내용을 발견하기가 매우 힘들며, 분식내용이 기업실패를 왜곡시키는데 큰 영향을 갖는다.

넷째, 우리나라의 기업운영상 세무회계 중심의 재무정보로 인하여 기업실상을 왜곡시키는 경향이 크다. 우리나라 경제개발계획의 추진 초기부터 기업의 창업과 지속적인 성장을 유도하기 위하여 다양한 세제상의 혜택을 기업에 제공해 왔으며, 기업들이 이를 수용하기 위해서는 세무회계 중심으로 재무제표를 작성하지 않을 수 없었다. 그로 인하여 재무제표의 분식이 정부의 유도 하에 관행화 되었다.

다섯째, 재무제표의 작성기준에 자의성이 많다. 기업회계기준으로 인정되는 회계처리방법과 절차 중에서 어떤 것을 선택할 것이며, 일단 선택된 것을 실제로 어떻게 할 것인가를 결정하는 데는 회계주체인 기업과 회계담당자의 자의적인 판단이 작용하게 된다. 이와 같은 회계의 자의성에 따른 회계결과는 큰 차이를 나타낼 수 있다.

여섯째, 재무제표상의 자산은 본연의 실질적인 가치를 나타내지 않는다. 고정자산의 감가상각 제도 및 자산 재평가 제도 등은 이를 잘 대변해 주고 있다.

일곱번째, 재무제표의 비교 가능성에도 상당한 한계가 있다. 따라서 각 회계연도 간, 기업간의 비교가 어렵다.

이와 같이 재무정보만으로 신용을 평가하는 데는 앞에서 지적한 바와 같이 상당한 한계와 문제점이 있다는 것을 알 수 있다. 즉 재무제표에 의한 소기업의 위험측정은 현실적으로 어려우며, 이는 금융기관이 소기업의 순수신용을 바탕으로 여신을 제공할 수 없는 근본적인 원인이 되고 있다.

(2) 소기업 신용평가에 있어 질적 요인의 중요성

재무정보를 토대로 한 양적 정보는 수량정보로서 객관화하기는 쉽지만 정보의 범위가 대단히 제한적이기 때문에 신용평가 및 기업도산의 예측에 한계를 지니고 있다. (손정식외 2인, 1992) 따라서 기업의 신용평가, 특히 소기업의 신용평가 시 비재무정보(non-financial information)의 중요성이 점차 부각되고 있다.

비재무정보는 대개 공시되지 않은 개별적, 특징적 요인들이 많이 포함하고 있어 그 중요성에도 불구하고 이에 대한 연구가 체계적으로 발전되지 못하였다. 그러나 최근에는 비재무정보(기업의 질적정보)를 활용한 기업의 신용평가가 미국, 일본 등에서 활발

히 연구되고 있으며, 우리나라의 금융기관 및 신용평가 전문회사에서도 부실기업에 대한 조사를 통해 비재무적 정보의 영향을 확인하고 있다.

외국의 연구사례를 살펴보면, Horrigan은 채무의 등급을 결정하는 모형의 정립에 있어서 독립변수로 재무정보 외에 비재무 정보의 하나인 채무변제의 우선순위를 가변수로 추가함으로써 모형의 예측력을 크게 높였다고 보고하고 있다.(J.O. Horrigan, 1965) 또한 Argenti역시 경영관리자의 관리양태, 회계제도의 양부 등 10개 변수를 점수화하여 도산 예측모형을 정립한 바 있다.(J. Argenti, 1977)

실제 금융기관의 기업의 신용평가, 대출 심사나 조사 등의 업무에서도 재무자료를 중심으로 각종 질적자료를 가미하여 실시되는 경우가 보다 일반화되어 가고 있다. 즉, 비재무 정보는 재무제표 정보의 객관화하기는 쉬우나 정보의 범위가 제한적이라는 약점을 보완하고 하고 있는 것이다.(민경선, 1989)

한편 최근 들어 상당한 성능을 갖춘 통계프로그램의 등장과 개인용 컴퓨터의 이용으로 비재무 정보에 대한 분석을 객관화할 수 있게 되었으며, 특히 질적요인, 즉 범주형 자료에 대한 분석기법이 소개되면서 향후 이에 대한 연구성과가 기대되고 있다.

III. 연구방법

본 연구에서는 소기업 신용평가모델을 개발하기 위한 연구방법으로 통계분석기법과 의사결정나무를 이용한다. 우선 소기업의 건전 및 부실을 예측하는데 유의한 영향을 주는 변수들은 단일변량 통계분석인 t-검정, Mann-Whitney-Wilcoxon검정, 적합성 검정을 위한 χ^2 -검정을 통해 추출해 낸다. 통계분석을 통해 유의성이 검정된 변수들은 의사결정나무의 후보 입력변수로 활용되어 의사결정나무를 통한 신용평가 모델 구축의 효율성을 높인다.

여기서는 분석에 이용된 각각의 통계기법들과 의사결정나무모형에 대해 자세히 살펴 보도록 하겠다.

A. 통계분석 기법

1. 두 독립 정규집단의 비교: t-검정

독립적인 두 모집단 평균차이의 유의성을 검증하는 통계기법으로 크게 Z-검정(Z-test)과 t-검정(t-test)으로 구분되는데 Z-검정은 모집단의 분산을 알고 있는 경우에 사용된다. 그러나 두 모집단의 분산을 알고 있는 경우는 드물기 때문에 보통 t-검정을 사용하게 된다.(정충영, 최이규, 2000)

각 표본의 관측값들은 서로 독립이며 동일한 정규분포를 따르고, 두 표본은 각각 독립이라 가정한다. 이때 모집단의 분산들은 동일하다고 가정하자.

$$\begin{aligned} X_1, X_2, \dots, X_m &\sim N(\mu_1, \sigma_1^2) \\ Y_1, Y_2, \dots, Y_m &\sim N(\mu_2, \sigma_2^2) \end{aligned}$$

귀무가설과 대립가설을 다음과 같이 설정하였을 때

$$H_0 : \mu_1 = \mu_2 \quad VS. \quad H_0 : \mu_1 \neq \mu_2$$

귀무가설이 옳다는 가정에서 검정통계량은 자유도 $m+n-2$ 의 t -분포에 따르게 된다.

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{S_p^2 \left(\frac{1}{m} + \frac{1}{n} \right)}}$$

여기서

$$S_p^2 = \frac{(m-1)S_1^2 + (n-1)S_2^2}{(m-1) + (n-1)}$$

는 합동분산(pooled variance)이라고 하며, 합동분산 통계량은 σ^2 의 추정량이다. S_1^2

과 S_2^2 는 표본 1과 표본 2의 표본분산이다. 가설을 검정하기 위하여는, t 값의 절대값

이 큰 경우 귀무가설을 기각하게 된다.

한편, 두 표본의 모분산의 동일성에 대한 검정은 다음과 같다. 즉, 다음 가설

$$H_0: \sigma_1^2 = \sigma_2^2 \text{ VS. } H_1: \sigma_1^2 \neq \sigma_2^2$$

에 대한 F 값과 자유도, 그리고 대응하는 p -값에 의해 귀무가설의 기각여부를 결정한다.

2. 비모수적 검정

앞 절의 t -검정은 두 독립집단의 정규분포를 가정할 때 사용되는 모수적 방법이다. 모수적 방법(parametric method)은 관측값이 어느 특정한 확률분포(정규분포, 이항분포 등)를 따른다고 전제할 후 그 분포의 모수(parameter)에 대한 검정을 실시하는 방법이라 할 수 있다. 반면, 비모수적방법(nonparametric method)은 관측값이 어느 특정한 확률분포를 따른다고 전제할 수 없는 경우에 실시하는 검정방법으로 모수에 대한 언급이 없으며 분포무관방법이라고 하기도 한다. 본 논문에서는 표본의 정규분포성 뿐만 아니라 비정규분포의 측면을 고려한 비모수적 방법도 검정에 사용하였다.

비모수적 방법의 특징은 다음과 같다. 첫째, 최소한의 가정을 전제로 하므로 가정이

만족하지 않음으로써 발생하는 오류를 줄일 수 있다. 둘째 대부분의 경우 관측값 보다는 상대적 순위를 이용하여 검정통계량을 유도하므로 순서척도로 관측된 자료의 검정에 유용하다. 셋째 계산과정이 단순하고 통계적 의미를 이해할 수 있으므로 통계에 대한 깊은 지식을 필요로 하지 않는다. (원태연, 정성원, 2002)

가. Mann-Whitney-Wilcoxon 순위합검정

독립인 두 집단의 평균의 차이를 비교할 때 모집단의 가정이 만족된다면 t-검정을 사용해야 보다 정확한 검정을 할 수 있다. 그러나 모집단의 가정이 만족되지 않거나 모집단의 분포에 대한 완화된 가정이 요구된다면 t-검정 대신에 비모수적 방법인 Mann-Whitney-Wilcoxon 검정을 사용할 수 있다. 이때 데이터는 순위 또는 구간으로 나타낼 수 있어야 한다. Mann-Whitney-Wilcoxon 검정과 같은 다른 비모수 검정 기법들은 데이터의 분포에 대하여 엄격한 가정을 필요로 하지 않기 때문에 일반적으로 모수 검정보다 검정능력이 떨어진다. (원태연, 정성원, 2002)

Mann-Whitney-Wilcoxon의 순위합검정은 두 모집단의 위치모수의 동일성을 검정하기 위한 분포무관 검정법이다. 두 모집단의 확률표본으로 된 혼합표본에서 관측값에 순위를 부여하고, 한 표본에 속하는 관측값의 순위의 합으로 검정통계량이 정의되며 그 절차는 다음과 같다. (송문섭, 박창순, 1989)

- (1) 두 모집단으로부터 각각 크기 m 과 n 인 확률표본을 이라 한다.

$$\begin{aligned} X_1, X_2, \dots, X_m \\ Y_1, Y_2, \dots, Y_n \end{aligned}$$

$m \geq n$ 이라 가정하며, $m < n$ 인 경우에는 X 와 Y 를 서로 바꾸어 생각한다.

- (2) X 와 Y 의 혼합표본에서 X_j 의 순위를 S_j 라고 한다.

- (3) Mann-Whitney-Wilcoxon 순위합 통계량

$$W = \sum_{j=1}^n S_j$$

즉, 통계량 W 는 X 표본에 부여된 순위의 합을 나타낸다.

(4) 귀무가설 $H_0 : \Delta = 0$ 하에서 Mann-Whitney-Wilcoxon 순위합 검정통계량 W

는 그 평균 $\frac{n(m+n+1)}{2}$ 에 대하여 대칭인 분포를 갖는다. 따라서

$$w(\alpha, m, n) - \frac{n(m+n+1)}{2} = \frac{n(m+n+1)}{2} - w(1-\alpha, m, n)$$

임을 알 수 있고 이 식으로부터 관계식

$$w(\alpha, m, n) - \frac{n(m+n+1)}{2} = \frac{n(m+n+1)}{2} - w(1-\alpha, m, n)$$

을 얻는다.

(5) 검정법: 유의수준 α 에서 $H_0 : \Delta = 0$ 일 때

$$W \geq w\left(\frac{\alpha}{2}, m, n\right) \text{ 또는 } W \leq w\left(1-\frac{\alpha}{2}, m, n\right)$$

이면 H_0 를 기각한다.

나. 도수자료에 대한 동질성 검정 (χ^2 -검정)

모집단이나 측정결과를 어떤 특성에 따라 분류시켜 도수로 주어지는 자료를 범주형 자료(categorical data)라고 하며, 이러한 자료를 이용한 통계적 추론방법을 범주형 자료의 분석(analysis of categorical data)이라 한다.

한편 자료를 두개의 속성(범주)에 따라 분류시켜 표로 만들어 놓은 것을 분할표라고 하며, 이 때 두 속성간에 관계가 있는지를 알아보기 위하여 χ^2 -검정을 이용한다.

(전명식, 1996) 여기서는 행(row)의 변수가 2개의 범주를 가지고 열(column)의 변수가 C개의 범주를 가지며 표본도수의 형태는 다음과 같다.

<표4> 2×c 분할표

목표변수 설명변수	범주1	범주2	범주C	합계
범주1	f_{11}	f_{12}	f_{1c}	$f_{1.}$
범주2	f_{21}	f_{22}	f_{2c}	$f_{2.}$
합계	$f_{.1}$	$f_{.2}$	$f_{.c}$	$f_{..}$

단, f_{ij} 는 범주(i,j)에 속하는 관찰도수, $i=1,2, j=1,2,\dots,c$

이 절에서 고려하는 두 집단간의 동질성은 각 행에 대한 열범주의 확률이 동일하다는 가설이 될 것이다. 즉 P_{ij} 를 i 번째 집단에 대한 다항분포의 j 번째 범주에 속할 확률이라 할 동질성 가정 $H_0: P_{1j} = P_{2j}, j=1,2,3,\dots, c$ 을 뜻한다.

이 동질성 가설에 대한 카이제곱 통계량은 두 종류가 있는데, 첫번째 것은 피어슨의 카이제곱 통계량(Pearson's χ^2 -statistic)으로

$$\chi^2 = \sum \frac{(f_{ij} - e_{ij}^e)}{e_{ij}}$$

$$\text{단, } e_{ij} = \frac{f_{i.}f_{.j}}{f_{..}}$$

이며, 둘째 것은 우도비 카이제곱 통계량(likelihood ratio χ^2 -statistic)이다. (허명회, 1992)

$$\chi^2 = 2 \sum f_{ij} \times \log_e \left(\frac{f_{ij}}{e_{ij}} \right)$$

이때 이런 χ^2 -통계량이 그의 귀무분포(χ^2_{c-1})에 대해 지나치게 크면 집단의 동질성을 기각하게 된다. 각 변수에 대한 두 기업군의 동질성은 χ^2 -적합도라는 관점에서

두 기업군 변수의 각 수준에 대한 도수를 조사하여 이에 기초한 통상적인 적합도 검정을 시도하는 것이다.

B. 의사결정나무(Decision Trees) 기법

1. 의사결정나무의 개념

의사결정나무는 의사결정규칙(decision rule)을 나무구조로 도표화하여 분류(classification)와 예측(prediction)을 목적으로 하는 어떤 경우에도 사용될 수 있는 기법이다.

의사결정나무는 ‘차원 축소 및 변수선택’, ‘교호효과 파악’, ‘범주의 병합’ 또는 ‘연속형 변수의 이산화’로 볼 때 탐색관계(exploration)와 모형화(modeling)라는 두 가지 특성을 가지고 있다고 할 수 있다. 즉 의사결정나무는 판별 분석(discrimination), 회귀분석(regression analysis)들과 같은 모수적(parametric) 모형을 분석하기 위해서 사전에 이상치(outlier)를 검색하거나 분석에 필요한 변수 또는 모형에 포함되어야 할 교호효과를 찾아내기 위해서 사용될 수도 있고 그 자체가 분류 또는 예측 모형으로 사용될 수도 있다. (2002, 김희진)

2. 의사결정나무의 구성

의사결정나무는 하나의 나무구조를 이루고 있으며 마디(node)라고 불리는 구성요소들로 이루어져 있다. 뿌리마디에서 어느 한 끝마디(leaf node)로 가는 길은 유일하여 그 길은 규칙으로 표현가능하며 서로 다른 끝마디들이 같은 분류(classification)를 만들 수도 있지만, 각 끝마디는 서로 다른 근거에 의해 그 분류를 만든다.

뿌리마디 (root node)	나무구조가 시작되는 마디로서 전체 자료로 이루어져 있다.
자식마디 (child node)	하나의 마디로부터 분리되어 나간 두개이상의 마디들을 의미한다.
부모마디	자식마디의 상위마디를 의미한다.

(parent node)	
중간마디 (terminal node)	각 나무줄기의 끝에 위치하고 있는 마디로써 잎(leaf)이라고도 하며, 이 끝 마디의 개수만큼 분류규칙이 생성된다.
가지 (branch)	하나의 마디로부터 끝마디까지 연결된 일련의 마디들을 의미하며, 이때 가지를 이루고 있는 마디의 개수는 깊이(depth)라고 한다. 가지 수와 깊이는 다양하다.

3. 의사결정나무의 장, 단점

가. 장점

첫째, 나무구조로 표현되어 해석이 용이하고, 어떤 입력변수가 목표변수를 설명하는 데 있어서 더 중요한지를 쉽게 파악할 수 있다. 둘째, 회귀분석이나 로지스틱 판별분석과 같은 모수적 모형은 모든 가능한 교호효과를 모형에 포함시키기 어렵지만, 의사결정나무는 유용한 입력변수나 교호효과 또는 비선형성을 자동적으로 찾아내는 알고리즘이다. 셋째, 의사결정나무는 선형성(linearity), 정규성(normality), 또는 등분성(equal variance) 등의 가정을 필요로 하지 않는 비모수적 방법이다. 넷째, 순서형 또는 연속형 변수는 단지 순위만(rank)만 분석에 영향을 주기 때문에 이상치에 민감하지 않다는 장점을 가지고 있다. (김지현,1998)

나. 단점

첫째, 연속형 변수를 비연속적인 값으로 취급하기 때문에 분리의 경계점 근방에서는 예측오류가 클 가능성이 있다. 둘째, 분석용 자료(training data)에만 의존하기 때문에 새로운 자료의 예측에서는 불안정할 가능성이 높다. 따라서 검증용 자료(test data)에 의한 교차타당성(cross validation) 평가나 가지치기에 의해서 안정성이 있는 의사결정 나무를 얻는 것이 좋다.

4. 의사결정나무분석의 알고리즘

의사결정나무분석의 가지치기 알고리즘은 다양하지만 가장 보편적인 것으로 CART

(Classification and Regression Tree)와 CHAID(Chi-squared Automatic Interaction Detection)이고 좀더 새로운 알고리즘은 C4.5이며 이것을 개선한 알고리즘 C5.0을 들 수 있다. 본 연구에서 사용하는 Clementine5.2에서는 CART와 C5.0 알고리즘을 지원하고 있다.

가. CHAID (Chi-squared Automation Interaction Detection)

CHAID (Chi-squared Automation Interaction Detection)는 1975년 J.A.Hartigan에 의해 소개된 알고리즘으로 Automatic Interaction Detection System(AID)에 기원을 두고 있으며 이것은 두 변수간의 통계적 관계를 찾은 것이다.

(1) 이산형 목표변수에 대한 분리기준

CHAID는 목표변수가 이산형일 때, Pearson 카이제곱 통계량 또는 우도비 카이제곱 통계량(likelihood chi-square statistic)을 분리기준으로 사용한다. 여기서 목표변수가 순서형 또는 사전 그룹화 된 연속형인 경우 우도비 카이제곱 통계량이 사용된다. 카이제곱 통계량은 관측도수(f_{ij})로 이루어진 $r \times c$ 분할표(contingency table)에서 계산된다.

<표5> $r \times c$ 분할표(contingency table)

목표변수 설명변수	범주1	범주2	범주C	합계
범주1	f_{11}	f_{12}	f_{1c}	$f_{1.}$
범주2	f_{21}	f_{22}	f_{2c}	$f_{2.}$
.....
범주r	f_{r1}	f_{r2}	f_{rc}	$f_{r.}$
합계	$f_{.1}$	$f_{.2}$	$f_{.c}$	$f_{..}$

위의 분할표로부터, 피어슨 카이제곱 통계량(Pearson χ^2 -statistic)은

$$\chi^2 = \sum \frac{(f_{ij} - e_{ij}^e)}{e_{ij}}$$

과 같이 정의 되고, 우도비 카이제곱 통계량(likelihood ratio χ^2 -statistic)은

$$\chi^2 = 2 \sum f_{ij} \times \log_e \left(\frac{f_{ij}}{e_{ij}} \right)$$

로 정의 된다. 이때 두 통계량의 자유도는 $(r-1)(c-1)$ 로서 동일하다. 여기서 e_{ij} 는

분포의 동일성 또는 독립성의 가설하에서 계산된 기대도수를 말하며 $e_{ij} = \frac{f_{i.}f_{.j}}{f_{..}}$ 과

같이 계산된다.

카이제곱 통계량이 자유도에 비해서 작다는 것은 예측변수의 각 범주에 따른 목표변수의 분포가 서로 동일하다는 것을 의미한다. 따라서 예측변수가 목표변수의 분류에 영향을 주지 않는다고 결론지을 수 있다. 자유도에 대한 카이제곱 통계량 값의 크고 작음은 p-값으로 표현될 수 있는데, 카이제곱 통계량 값이 자유도에 비해서 작으며 p-값은 커지게 된다. 결국 분리기준을 카이제곱 통계량 값으로 한다는 것은, p-값이 가장 작은 예측변수와 그 때의 최적분리에 의해서 자식마디를 형성시킨다는 것을 의미한다.

각 예측변수에 대한 최적분리 탐색은 우선 분할표로부터 카이제곱 값을 계산한다. 그리고 예측변수의 측도에 따라 가능한 모든 경우에 대해서 가장 병합될 가능성이 큰 두 범주를 찾아내 이 경우에 대해서 병합에 대한 탐색을 수행한다. 여기서 병합은 어느 두 범주에 의해서 계산된 p-값이 지정된 유의수준보다 크다면, 두 범주에 따른 목표변수의 분포가 동일하다는 것을 의미하므로 이루어지는 것이다. 그리하여 $r \times c$ 분할표로부터 시작하여 더 이상 병합할 두 범주가 없을 때까지 탐색 후 최종적으로 $d \times c$ ($d \leq c$) 분할표를 얻는다.

p-값에 대한 Bonferroni 조정

분산분석에 있어서 다중비교 문제와 유사하게 각 예측변수에 대하여 최종적인 병합에 의해서 얻어진 p-값은 정확한 p-값보다 작게 된다. 따라서 Bonferroni계수를 p-

값에 곱하여 근사적으로 보다 정확한 p-값을 얻는 방법을 이용하는 경우가 발생하게 된다.

Bonferroni계수는 순서형 예측변수에 대해서는 이항계수(Binomial Coefficient)로부터

$$\beta_n = {}_{c-1}C_{r-1} \quad (c: \text{목표변수의 범주}, r: \text{설명변수의 범주})$$

와 같이 계산되고, 명목형 변수에 대해서는

$$\beta_n = \sum_{i=0}^{r-1} (-1)^i \frac{i(r-1)^c}{i!(r-1)!} \quad (c: \text{목표변수의 범주}, r: \text{설명변수의 범주})$$

와 같이 계산된다.

(2) 연속형 목표변수에 대한 분리기준

CHAID는 목표변수가 연속형인 경우에는 두 개 이상의 그룹에 대해서 평균치 차를 검정하는 분산분석(ANOVA table)의 F통계량을 분리기준으로 이용한다. F통계량은 자유도(r-1, n-r)인 F분포를 따르는 것으로 알려져 있다(r은 예측변수의 범주, n은 전체 관측치 수). F통계량이 자유도에 비해서 매우 작다는 것은 예측변수의 각 범주에 따른 목표변수의 평균치 차가 존재하지 않는다는 것을 의미하며, 이것은 예측변수가 목표변수의 예측에 영향을 주지 않는다고 결론지을 수 있다. 카이제곱 통계량과 마찬가지로 자유도에 대한 F-통계량의 크고 작음은 p-값으로 표현될 수 있는데 F통계량이 자유도에 비해서 작으면 p-값은 커지게 된다. 이와 같이 계산된 F통계량의 p-값을 기준으로 명목형 목표변수인 경우와 유사하게 병합과 분리를 계속하여, p-값이 가장 작은 예측변수와 그때의 최적분리에 의해서 자식마디가 형성된다.

나. CART

L. Breiman Friedman Olshen Stone(1984)에 의해 고안된 CART(Classification And Regression Tree)는 설명변수들의 특성에 따라 자료들을 이진분할하여 두 개의 하위 노드를 생산, 반응변수의 값이 유사한 부분집합으로 만드는 방법이다. 즉 노드의 불순도(impurity)를 최소화 시키는 것이 목적으로 이산형 목표변수인 경우에 적용되는 지니지수(Gini Index)와 연속형 목표변수인 경우 적용되는 분산의 감소량을 이용하여 이진분류(binary split)를 수행하는 알고리즘이다.

(1) 목표변수가 이산형인 경우의 분리기준

ㄱ. 지니지수(Gini Index)

앞에서 설명한 카이제곱 통계량과 마찬가지로 지니지수도 불순도(impurity)를 측정하는 하나의 지수이다. 각 마디에 속하는 개체를 그 마디에서 도수가 가장 많은 목표변수의 한 범주에만 모두 할당하는 분류규칙을 고려해 볼 때, 임의의 한 개체가 목표변수의 i 번째 범주로부터 추출되었고, 그 개체를 목표변수의 j 번째 범주에 속한다고 오분류(misclassification)할 확률을 모두 더하여

$$G = \sum_{j=1}^c \sum_{i \neq j} P(i)p(j)$$

를 얻을 수 있고, 위와 같은 분류 규칙 하에서 오분류 확률의 추정치(estimate)가 된다. 여기서 c 는 목표변수의 범주수를 말한다.

일반적으로 CART는 이산형 목표변수에 대해서는 지니지수(Gini index)를 분리기준으로 사용한다. 지니지수는 각 마디에서의 불순도(impurity) 또는 다양도(diversity)를 재는 측도 중의 하나로써

$$G = \sum_{j=1}^c P(i)(1 - P(j)) = 1 - \sum_{j=1}^c P(j)^2 = 1 - \sum_{j=1}^c (n_j / n)^2$$

와 같이 표현될 수 있다. 여기에서 n 은 그 마디에 포함되어 있는 관찰치 수를 말하고, n_j 는 목표변수의 j 번째 범주에 속하는 관찰치 수를 말한다. 지니지수는 n 개의 원소 중에서 임의로 두개를 추출하였을 때 추출된 두개가 서로 다른 그룹에 속해있을 확률을 의미하여 Simpson의 다양도 지수(diversity index)로도 알려져 있다. 목표변수의 범주가 2개인 경우에는 지니지수는 다음과 같이 표현될 수 있으며,

$$G = 2P(1)P(2) = 2\left(\frac{n_1}{n}\right)\left(\frac{n_2}{n}\right)$$

이는 카이제곱 통계량을 사용하는 것과 같은 결과를 갖는다.

CART는 이 지니지수를 가장 감소시켜주는 예측변수와 그 변수의 최적분리를 자식마디로 선택하는데, 지니계수의 감소량은 다음과 같은 계산된다.

$$\Delta G = G - \frac{n_L}{n} G_L - \frac{n_R}{n} G_R$$

(단, n =관측치 수, n_R, n_L =각각 자식마디의 관측치 수,

G_L = left node에서의 불순도, G_R = right node에서의 불순도)

즉, 자식마디로 분리 되었을 때의 불순도가 가장 작도록(순수도가 가장 크도록) 자식

마디를 형성하는 것이다.

$$P(L)G_L + P(R)G_R = \frac{n_L}{n}G_L + \frac{n_R}{n}G_R$$

(2) 목표변수가 연속형인 경우의 분리기준

목표변수가 연속형일 때는 각 마디의 다양도의 측도로써 다음과 같은 분산 (variance)를 고려할 수 있다.

$$V = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$$

이는 마디의 목표변수의 평균을 그 마디에 속하는 모든 개체의 예측값으로 사용하는 것을 고려한다면 예측오차를 최소화 하는 것과 동일하다 할 수 있다.

명목형 목표변수인 경우와 마찬가지로 분산의 감소량을 최대화 하는 것은 다음과 같은 기준

$$\Delta V = V - \frac{n_L}{n}V_L - \frac{n_R}{n}V_R$$

을 최대화 하는 것과 같다.

다. C5.0

(1) C4.5

C4.5는 호주의 J.Ross Quinlan에 의해 정립된 대표적인 의사결정나무 알고리즘으로, 초기에는 ID3 (Iterative Dichotomiser 3)의 이름으로 하여 1986년에 기계학습분야에서 적용하여 발표되었고 이 분야에서 매우 효과적인 이론으로 소개되었다. C4.5의 특징은 순서형 예측변수에 대하여는 이진분리를 시행하고, 범주형 예측변수에 대해서는 다원분리를 시행하는 것이다. C4.5에서 사용하는 분리측도는 엔트로피의 감소로 정의되기도 하는 이득비율이다. 그러나 이득에 기초한 기준은 범주의 개수가 많은 변수로의 편의가 심각하여 '(엔트로피의 감소)/(분리의 엔트로피)'로 정의되는 이득비율을 분리측도로 제안하였다.

ID3는 정보이론(Information Theory)에 Shannon(1948)에 따른 엔트로피(entropy) 개념을 이용하여 가장 간단한 의사결정나무를 구하고 있다. 일반적으로 엔트로피는 무질서도를 나타내는 수치이다. 즉 일반적으로 ID3에 입력된 입력사례 집합에 대해 적합한 사례와 적합하지 않은 사례가 혼합되어 있으므로 어떤 속성이 어떻게 사례를 분류

하는지에 대한 정보가 없기 때문에 엔트로피가 아주 높은 상태이다. 그러나, 일단 결정나무가 학습된 상태에서는 각각의 말단 마디가 하나의 등급으로만 결정이 되므로 무질서도는 0이 된다. 그러므로 ID3이 사례를 끝까지 분류한 상태에서는 엔트로피가 0이 되는 상태를 의미하므로, 학습사례를 가장 잘 분류하는 속성은 정보량의 감소량을 최대화하는 것이다. 따라서 각 속성들이 분류기준으로 선택될 경우 정보량의 감소량을 계산한다. 다음으로 가장 식별력이 높은 속성이 분류기준으로 선택되면, 주어진 속성값의 종류 만큼 가치를 만든다. 각각의 가치에 해당하는 값에 따라 사례를 분할하고, 각각의 가치에서 지금까지의 과정을 반복한다. 더 이상의 정보량의 감소가 없다면 분할을 멈춘다.

즉, ID3는 Entropy Function에 근거한 정보이론 접근방식(Information Theoretic Approach)를 이용하여, Training events를 변수(attributes)에 대해 올바르게 분류하기 위해 Top-down divide-and-conquer 방식으로 의사결정 나무를 구성하도록 하는 알고리즘이다.

(2) 정보의 정의

정보이론(Information Theory: Claude Shannon, 1948)에 의하여 정보를 전달하는 최소단위는 “예” 혹은 “아니오”에 의해 연속적으로 산출되어진 조합에 의해 전달을 할 때 가장 최소의 용량을 필요로 한다. 이에 대한 정보전달을 할 때 필요한 최소 단위는 bit이고 이것은 $\log_2(\text{데이터의 수})$ 로 표현되어 진다. 따라서 2진법으로 정보를 표현하는데 필요한 비트수를 정보량(I)으로 표현하도록 한다.

위의 정보이론을 바탕으로 n개의 동등한 확률을 갖는 메시지가 있다고 가정했을 때, 각각의 확률은 각각 $p=1/n$ 이 된다. 따라서 그 때의 각각의 메시지가 전달되어지는 정보는 다음과 같은 수식으로 설명할 수 있다.

$$I = -\log_2(p) = \log_2(n)$$

예를 들어 각기 다른 16개의 메시지가 있다고 가정하였을 때, $\log_2(16) = 4$ 가 된다. 즉, 정보량은 4가 된다. 이것의 의미는 16개 각각의 메시지를 확인하는데 4 bits가 필요하다는 의미이다.

(3) 이익표준(Gain Criterion)

이익표준은 어떤 변수가 목표변수에 대하여 얼마만큼의 정보력을 가지고 있는가를 측정하는데 사용되며 다음과 같이 정의된다.

$$\text{Gain}(X, T) = \text{Info}(T) - \text{Info}(X, T)$$

이 식은 변수 X의 값이 얻어진 후의 T의 요소를 구별하는데 필요한 정보와 단순히 T의 요소를 구별하는데 필요한 정보의 차로 설명할 수 있다. 이른바 우리는 이것을 변수 X에 따른 정보의 획득(The gain information due to attribute X)이라 한다.

Gain Criterion을 사용하는 의사결정 규칙은 정보를 가장 많이 가지고 있는 변수의 산출에 기반이 되는 일부 데이터를 선택하여 분할하게 된다. 이렇듯 Gain Criterion은 정보력이라는 분야에 대해서는 상당히 훌륭한 결과를 나타내지만, 그 책정기준이 큰 수의 결과물을 갖는 분할영역을 선호한다는 점은 단점이라 할 수 있다. 따라서 이것을 보완하기 위해 이익비율표준(Gain Ratios Criterion)을 사용하게 된다.

(4) 이익비율표준(Gain Ratios Criterion)

이익 표준(Gain Criterion)은 큰 수의 값을 갖는 변수를 선호한다. 예를 들어, 각 레코드가 별개의 값을 갖는 변수 D가 있다고 할 때, Info(D,T)는 0이 된다. 따라서 이 경우, Gain(D,T)는 최대가 된다. 하지만 이 경우가 최대의 정보력을 갖는다고 할 수 없을 것이다. 따라서 우리는 비율적 Gain을 고려하여야 되는데, 이를 위하여 이익 비율 표준(Gain Ratios Criterion)을 사용하여야 한다.

$$GainRatio(D,T) = \frac{Gain(D,T)}{SplitInfo(D,T)}$$

$$: SplitInfo(D,T) = I\left(\frac{|T_1|}{|T|}, \frac{|T_2|}{|T|}, \Lambda, \frac{|T_n|}{|T|}\right)$$

(T_1, T_2, K, T_n) : 변수 D의 값에 의해 산출되어진 T의 분할 영역

여기서, SplitInfo(D,T)는 범주형 변수 D의 값에 근거한 데이터 T의 분리된 정보를 말한다. 즉, 이익(Gain)은 분류 및 군집과 관련된 정보를 측정하는 것에 반하여 SplitInfo(D,T)는 n개의 결과에 속한 데이터를 분할하는데 산출되어진 잠재적 정보를 일컫는다. 따라서 SplitInfo(D,T)는 다음과 같은 수식으로 나타낼 수 있다.

$$I\left(\frac{|T_1|}{|T|}, \frac{|T_2|}{|T|}, \dots, \frac{|T_n|}{|T|}\right)$$

여기서 $\{T_1, T_2, \dots, T_n\}$ 은 변수 D의 값에 의해 산출되어지는 T의 분할영역을 의

미한다. 즉, 이익(Gain)은 분류 및 군집과 관련된 정보를 측정하는 것에 반하여 SplitInfo(D,T)는 n개의 결과에 속한 데이터를 분할하는데 산출되어진 잠재적 정보를 일컫는다.

(5) C5.0으로의 확장

C5.0에서의 분할기준은 이익비율표준(Gain Ratios Criterion)을 사용하여 변수의 결과들에 대하여 정보에 대한 이익비율(Gain Criterion)이 최대가 되는 점에서 데이터의 분할을 선택한다. 즉 정보력이 절대적 기준에 의해 확실히 크거나, 모든 자료에 대한 테스트 결과 이익(Gain)의 평균보다 적어도 큰 경우에 강제적으로 최대화를 하게 된다. 이러한 강제적인 최대화는 분할 시 최소한의 차이나 분할 정보가 작은 경우와 같은 불안정한 이익비율(Gain Criterion)의 경우는 무시하게 된다.

C5.0의 가지노드가 부모노드로부터 가지쳐지는 것을 결정하는 방법은 부나무(Sub-tree)에 대한 예측오차들과 잎(leaf)에 대한 예측오차를 비교하여 가지치기를 결정하게 된다. 즉, 가지쳐진 노드와 그렇지 않은 가지의 예측오차를 비교하여 결정하게 된다.

IV. 실험설계

A. 데이터 수집방법 및 표본기업

본 연구의 연구대상 모집단(target population)은 우리나라 소기업이며, 표본을 추출하기 위하여 규정된 모집단인 조사모집단(sample population)은 C은행과 현재 거래 중이거나 과거에 거래 경험이 있는 사업자 등록증을 보유한 총자산 10억 이하의 중소기업인 또는 개인사업자라 규정한다.(단, 2002년 현재, 정상영업기간이 1년 미만인 사업자는 제외한다.)

한국신용정보(주)는 이들 총 79,784개의 고객들 중 2,581개(건전: 1,293개, 부실: 1,287개)의 고객들을 임의로 추출하여 소기업 신용평가를 위한 고객개요정보 수집을 목적으로 설문을 실시하였다. 본 설문에 이용된 설문 문항들은 한국신용정보(주)와 현재 C은행에서 신용평가 심사역으로 활동하고 있는 전문가들이 토의를 통해 소기업 신용평가 시에 주요하게 취급하는 항목들을 위주로 선정하였다.

본 설문은 2001년 12월1일부터 1월1일에 걸쳐 약 1개월 동안 실시하였으며 표본으로 선택된 소기업 고객들에 대해서는 직접방문이나 전화면담을 통해 설문을 하는 형식으로 진행되었다. 수집된 데이터 중 본 연구에서 최종적으로 실험에 이용된 표본기업은 설문에 대해 무응답이 없는 총 789개(건전: 414개, 부실: 375개)의 기업이다.

B. 연구변수 및 연구방법의 설계

1. 연구변수

본 연구에서 분석대상으로 선정한 변수는 C은행 소기업 고객을 대상으로 한 설문 문항이다. 또한 각각의 문항들은 데이터 수집방법에서 밝힌 바와 같이 C은행의 신용평가 전문 심사역과 한국신용평가(주)의 신용평가 전문가들이 토의를 거쳐 소기업 신용평가와 연관된 질적 요인 및 선행 연구에서 그 유용성이 확인된 문항들을 우선으로 선정하였다.

구체적으로는 소기업의 건전/ 부실 여부를 판단하는데 영향을 주는 비재무적 변수로

서 ‘기업개요정보’, ‘기업현황정보’, ‘대표자 개요정보’, ‘대표자 신상/ 재산에 대한 정보’와 같은 4가지 카테고리로 항목들을 분류하였고 총 35개의 문항으로 이루어진 설문지를 구성하였다. 그러나 본 연구에서는 이들 35개의 항목 중 업종별 특이문항과 결측값을 35%이상 포함한 문항들은 제외시켰고 나머지 28개의 변수를 중심으로 조사를 실시하였다.

각각의 부실기업이 갖는 독자적인 특성은 본 논문에서 다루고자 하는 28가지 요인들 이외에도 다양하게 있다. 그러나 이러한 독립된 또는 공통된 많은 요인들을 찾아내어 관리하고, 비교 평가하는 일은 제반 비용이 과다하게 소요되어 현실적으로 불가능하며, 때로는 유의한 결과를 찾아내는 데 기여하지 않을 수도 있다. 따라서, 본 논문에서는 신용평가 전문가들에 의해 선정된 변수들을 중심으로, 중소기업의 신용평가 모델 구축을 위한 변수로서 활용 분석해 보고자 한다. 위의 과정을 통해 선정된 분석대상 변수 및 세부내용은 다음의 표와 같다.

<표6> 분석대상 변수 및 세부내용

	Field Name	내용	입력 코드	세부내용	변수유형
1	ID	ID #		아이디	
2	cmp_cd	KEY		기업코드	
3	G/B	우/불량	1	건전	명목척도
			2	불량	
4	X1	법인,개인	1	법인	명목척도
			2	개인	
5	X2	업종	1	제조	명목척도
			2	건설	
			3	도매	
			4	소매	
			5	음식 및 숙박업	
			6	임대	
			7	전문인적서비스	
			8	기타 물적 서비스	

6	X3	동업여부	1	동업	명목척도
			2	동업안함	
7	X4	동업계종사기간 (경영자경력)	1	그외기업	구간척도
			2	동업종 종사년수 3년이상	
			3	동업종 종사년수 5년이상	
			4	동업종 종사년수 7년이상	
			5	동업종 종사년수 10년이상	
8	X5	사업경력	1	그외기업	구간척도
			2	대표자로서 동업종 종사년수 3년이상	
			3	대표자로서 동업종 종사년수 5년이상	
			4	대표자로서 동업종 종사년수 7년이상	
			5	대표자로서 동업종 종사년수 10년이상	
9	X6	설립년(업력)	1	그외기업	구간척도
			2	설립후 3년이상	
			3	설립후 5년이상	
			4	설립후 7년이상	
			5	설립후 10년이상	
9	X7	정규직 수	제조업일 경우(50인이하)		구간척도
			1	10인이하	
			2	30인이하	
			3	50인이하	
			건설업일 경우(30인이하)		
			1	10인이하	
			2	20인이하	
			3	30인이하	
			그외 업종일 경우(10인이하)		

			1	3人以下	
			2	6人以下	
			3	10人以下	
10	X8	매출액	정수 범위	(만원)	
11	X9	총차입금 대 매출액비율 (=총대출액/매출액)	1	5%미만	구간척도
			2	5%이상	
			3	15%이상	
			4	30%이상	
			5	40%이상	
12	X10	사업장위치	1	주거지역	명목척도
			2	상업지역	
			3	유흥가	
			4	공단지역	
			5	도서벽지/기타	
13	X11	사업장 규모	정수	(평)	등비척도
14	X12	사업장 소유주	1	본인	명목척도
			2	부모	
			3	배우자	
			4	형제	
			5	친척	
			6	타인	
			7	사업장 없음	
15	X13	계열,관계사의 유무	1	있음	명목척도
			2	없음	
16	X14	배우자 직업	1	사업에 같이 참여	명목척도
			2	가사	
			3	회사원/공무원	
			4	다른 사업을 하고 있음	

			5	전문직	
			6	미혼	
			7	기타	
17	X15	주택 평수	정수	(평)	비율척도
18	X16	주택 소유자	1	본인	명목척도
			2	부모	
			3	배우자	
			4	형제	
			5	친인척	
			6	타인	
19	X17	주거형태	1	단독주택	명목척도
			2	아파트	
			3	빌라(다세대, 연립, 다가구)	
			4	사업장주거	
			5	상가주택	
20	X18	주택 소유기간(일)	정수	(일)	등비척도
21	X19	주택 거주기간(일)	정수	(일)	등비척도
22	X20	주택시가(만원)	정수	(만원)	등비척도
23	X21	자가용 배기량(CC)	1	소형(1500CC이하)	구간척도
			2	중형(2000CC이하)	
			3	대형(그 외의 경우)	
24	X22	당좌거래은행 개수	1	해당사항 없음	구간척도
			2	1개금융기관	
			3	2개금융기관	
			4	3개금융기관	
			5	4개금융기관	
			6	5개금융기관 이상	
25	X23	은행외 거래 개수	1	해당사항없음	구간척도
			2	1개금융기관	

			3	2개금융기관	
			4	3개금융기관	
			5	4개금융기간	
			6	5개금융기관 이상	
26	X24	대출액	정수	(만원)	등비척도
27	X25	연 소득액	정수	(만원)	등비척도
28	X26	연 재산세	정수	(만원)	등비척도
29	X27	과세 종류	1	일반과세자	명목척도
			2	간이과세자	
30	X28	부가세	정수	(만원)	등비척도

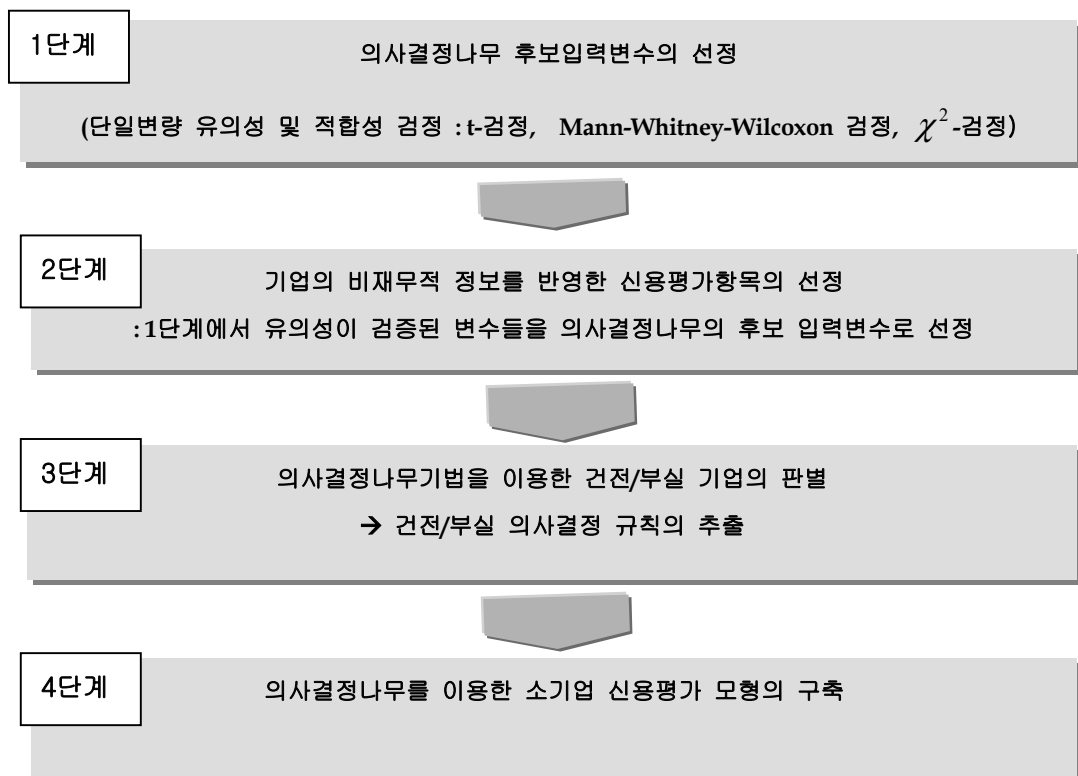
2. 연구방법의 설계

가. 입력변수의 선정

본 연구에서는 의사결정나무를 이용하여 소기업 신용평가모형을 개발하기에 앞서 소기업의 건전 및 부실을 예측하는데 유의한 영향을 주는 변수들을 추출하여 입력변수로 활용함으로써 의사결정나무를 통한 신용평가 모형 구축의 효율성을 높이하고자 하였다.

각 변수별 건전, 부실기업간 평균의 차이를 비교 분석하기 위해 본 연구에서는 Mann-Whitney-Wilcoxon 검정, t-검정, 집단간 동질성을 검정하는데 쓰이는 기법인

χ^2 적합성 검정과 같은 다양한 단일변량 검정통계기법을 이용하였다. 이후 단일변량의 검정에서 선정된 유의한 항목들은 이후 의사결정나무의 후보입력변수로서 건전/부실 기업의 판별 분석에 활용하였다. 이와 같은 모형구축과정은 아래의 <그림2>와 같이 도식화 할 수 있다.



<그림2> 소기업 신용평가 모형구축 과정

나. 의사결정나무의 C5.0 알고리즘을 이용한 소기업 신용평가 모형의 구축

본 연구에서는 앞에서 제시된 단일변량 분석기법들을 통해 선정된 변수들을 의사결정나무 모형의 입력변수들로 이용하여 소기업 신용평가 모형을 구축하였다. 총 789개의 데이터를 실험에 사용하였으며, 의사결정나무 모형을 구축하는데 쓰이는 데이터(609개)와 구축된 모형을 검증하기 위한 약 20%의 검증용 자료(180개-건전:90개, 부실:90개)로 나누어 실험을 실시하였다. 검증용 데이터는 비계통 임의 추출을 통해 90개(건전 45개, 부실 45개)씩 총 4개의 다른 data set을 만들어 검증을 실시 하였다. 또한 모형의 일반성을 유지하기 위해 75% severity 가지치기와 한 Parent 당 최소 10개 이상의 데이터가 포함되도록 하는 정지 규칙으로 고정시킨 후 실험을 실시 하였다. 본 연구에서 의사결정나무모형을 구축하는데 활용된 알고리즘은 C5.0이며 사용된 소프트웨어는 Clementine5.2이다.

<표7> 의사결정나무에서 사용된 훈련용, 검증용 데이터

구분	추출방식	표본			
		건전(1)	부실(2)	합계	비율
훈련용 자료	임의추출	324	285	609	약 80%
검증용 자료1	비계통 임의추출	45	45	90	약10%
검증용 자료2		45	45	90	약10%
검증용 자료3		45	45	90	약10%
검증용 자료4		45	45	90	약10%

V. 실험결과

A. 입력변수의 선정- 단일변량 분석

1. 건전기업과 부실 기업의 평균 차이 비교 : t-검정

t-검정은 앞에서 기술한 것과 같이 정규분포를 가정한 독립적인 두 모집단의 평균차이의 유의성을 검증하는 통계기법이다. 본 연구에서는 건전기업과 부실기업의 평균의 차이에 대해 각각의 변수들의 유의성을 분석하는데 이용하였는데, 모집단의 정규분포를 가정할 수 없는 명목변수의 경우는 t-검정에서 제외하였다.

변수별 t-검정 분석결과는 아래의 <표8>와 같이 정리해 볼 수 있다.

<표8> t-검정 결과 및 유의한 변수

변수명	설명	T	P값	sig.
X1	법인/개인	2.323	0.021	**
X2	업종	2.526	0.012	**
X3	동업			
X4	경영자의 동업계 종사기간	6.188	0	***
X5	경영자사업경력	2.291	0.022	**
X6	업력(구간)			
X7	종업원수(범주화)	0.911	0.363	
X8	매출액	2.65	0.008	***
X9	총차입금 대 매출액비율	-5.905	0	***
X10	시업장 위치			
X11	사업장규모(평수)	1.781	0.075	*
X12	사업장소유주			
X13	계열사/관계사유무			
X14	배우자 직업			

X15	주택규모(평수)	2.604	0.009	***
X16	주택소유자			
X17	주거형태			
X18	주택소유기간	3.758	0	***
X19	주택거주기간	2.6	0.01	***
X20	주택가시가	-1.483	0.139	
X21	자가용배기량(범주화)			
X22	당좌거래 은행개수			
X23	은행외거래 금융기관개수			
X24	총대출액(만원)	-1.262	0.208	
X25	연소득액(만원)	1.464	0.144	
X26	연재산세(만원)	2.388	0.017	**
X27	과세종류			
X28	부가세(만원)	1.711	0.088	*

※***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$

따라서 본 분석 결과로부터 법인/개인(X1), 업종(X2), 경영자의 동업계 종사기간(X4), 경영자의 사업경력(X5), 매출액(X8), 총차입금 대 매출액 비율(X9), 사업장규모(X11), 대표자 소유의 주택규모(X15), 주택소유기간(X18), 주택거주기간(X19), 연재산세(X26), 부과세(X28) 등의 변수들은 부실기업과 건전기업을 구분해 주는 유의한 변수임을 알 수 있다. 그 중 총차입금 대 매출액 비율의 t값을 살펴보면 부실기업의 평균이 건전기업보다 유의할 정도로 매우 차이가 나는 것을 알 수 있는데, 이는 건전기업보다 부실기업 일 경우 매출액에 대한 차입금 비율이 높음을 시사하고 있다.

2. Mann-Whitney-Wilcoxon 검정

모집단의 정규분포를 가정할 수 없거나, 모집단의 분포를 정규분포로 가정 할 수 없는 경우 t-검정 대신 비모수적 방법인 Mann-Whitney-Wilcoxon검정을 이용할 수 있다. 본 연구에서 사용된 변수들은 범주화, 구간화와 같은 방법을 통해서 범주화 되도록 처리하거나 자체가 명목척도인 변수가 다수 있었다. 이로 인해, 사용된 변수들의 모집단의 분포가 정확히 정규분포를 따르지 않거나 정규분포와 유사하긴 하나 분포에

대해 다소 완화된 가정을 해야 하는 경우가 많음을 알 수 있다. 따라서 본 연구에서는 건전, 부실기업의 평균의 차이에 대한 변수들의 유의성 검정에 Mann-Whitney-Wilcoxon검정을 사용하여 단일변량 검정을 수행해 보았으며 그 결과는 아래의 <표9>와 같다.

<표9> Mann-Whitney-Wilcoxon 검정결과

변수명	설 명	Mann-Whitney U	Wilcoxon W	P값	Sig.
X1	법인/개인	42453	83208	0.021	**
X2	업종	40867.5	81622.5	0.013	**
X3	동업	46020	86775	0.856	
X4	경영자의 동업계 종사기간(구간)				
X5	경영자사업경력(구간)				
X6	업력(구간)	43518	84273	0.199	
X7	종업원수(범주화)	44493	85248	0.435	
X8	매출액	43948	84703	0.305	
X9	총차입금대매출액비율	33013	85663	0	***
X10	사업장위치	43527.5	84282.5	0.165	
X11	사업장규모(평수)	45248.5	97898.5	0.67	
X12	사업장소유주	41637	94287	0.011	**
X13	계열사/관계사유무	43123.5	83878.5	0.002	***
X14	배우자 직업	44691	97341	0.459	
X15	주택평수(평수)	36335.5	77090.5	0	***
X16	주택소유자	43961.5	96611.5	0.104	
X17	주거형태	42163	94813	0.039	**
X18	주택소유기간	38914.5	79669.5	0.001	***
X19	주택거주기간	40506.5	81261.5	0.009	***

X20	주택가시가	45463.5	98113.5	0.138	
X21	자가용배기량(범주화)	44321	85076	0.355	
X22	당좌거래은행 개수	42848	83603	0.11	
X23	은행외거래금융기관개수	42587.5	95237.5	0.043	**
X24	총대출액(만원)	39168	91818	0.001	***
X25	연소득액(만원)	45330	97980	0.698	
X26	연재산세(만원)	31067.5	71822.5	0	***
X27	과세종류	45675	86430	0.741	
X28	부가세(만원)	44170	84925	0.356	

※***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$

검증결과, 법인/개인(X1), 업종(X2), 경영자의 동업계 종사기간(X4), 경영자의 사업경력(X5), 총차입금 대 매출액 비율(X9), 사업장 소유주(X12), 계열사/관계사유무(X13), 주택평수(X15), 주거형태(X17), 주택 소유기간(X18), 거주기간(X19), 은행 외 거래 금융기관개수(X23), 총대출액(X24), 연재산세(X26)의 변수들이 부도기업과 건전기업에 따라 구별되는 유의한 변수로 분석되었다.

3. 카이제곱 적합도 검정

범주화 변수나 명목변수에 대해서 건전, 부실 두 기업군의 동질성 검정은 χ^2 적합도 검정을 통해 실시하였다. χ^2 적합도 검정결과는 아래의 <표10>와 같다.

<표10> χ^2 적합도 검정결과

변수명	설명	피어슨 카이제곱	p값	Sig
X1	법인/개인	0.535	0.464	
X2	업종	21.099	0.004	***

X3	동업	0.061	0.805	
X4	경영자의 동업계 종사기간(구간)	14.484	0.006	***
X5	경영자사업경력(구간)	30.533	0	***
X6	업력(구간)	6.417	0.17	
X7	종업원수(범주화)	34.34	0	***
X10	사업장위치	4.995	0.288	
X12	사업장소유주	12.921	0.024	**
X13	계열사/관계사유무	6.459	0.011	**
X14	배우자 직업	28.692	0	***
X16	주택소유자	34.048	0	***
X17	주거형태	6.874	0.143	
X21	자가용배기량(범주화)	1.791	0.408	
X22	당좌거래은행 개수	9.422	0.051	*
X23	은행외거래금융기관개수	3.453	0.63	
X27	과세종류	0.151	0.698	

※***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$

적합성 검정결과, 업종(X2), 경영자의 동업계 종사기간(X4), 경영자의 사업경력(X5), 종업원수(X7), 사업장 소유주(X12), 관계사/계열사의 유무(X13), 배우자의 직업(X14), 주택 소유자(X16), 당좌거래 은행 수(X22)의 변수들은 건전기업과 부실기업에 대해 동질적이지 않다는 결과가 도출되었다.

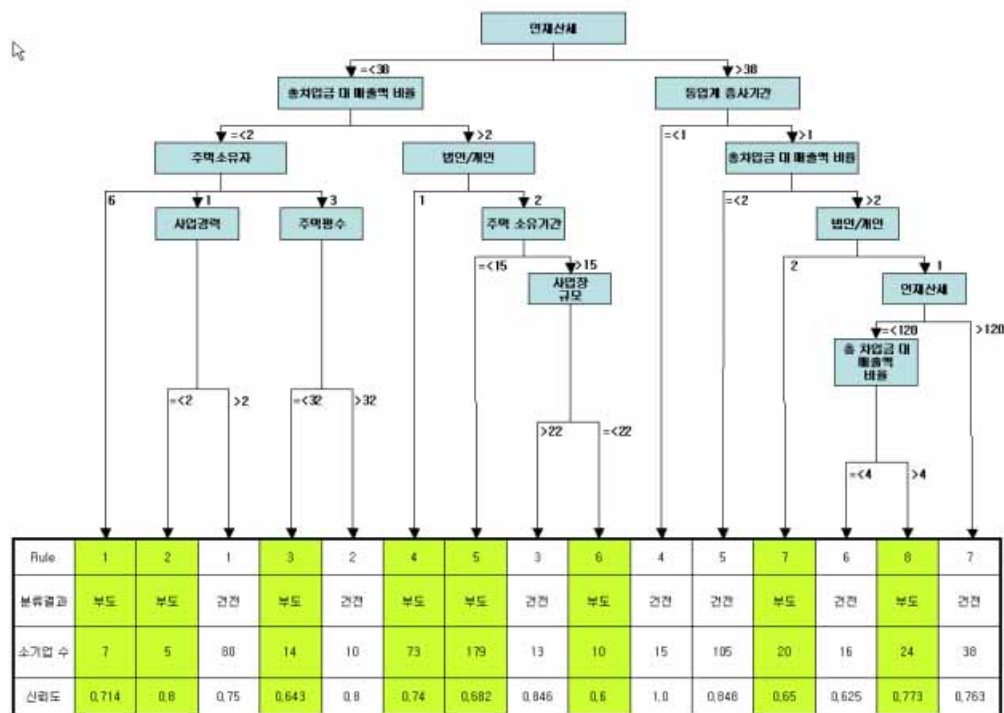
B. 의사결정나무에 의한 신용평가 모형

1. 의사결정나무의 구축 - Pruning전

앞에서 실시한 단일변량 테스트(t검정, Mann-whitney wilcoxon검정, χ^2 적합도 검정) 결과 선정된 변수들(X1: 법인/개인여부, X2: 업종, X4: 경영자의 동업계종사기간, X5: 대표자로서 동업계 사업경력, X8:매출액, X9: 총차입금 대 매출액비율, X11: 사업장규모, X12:사업장 소유주, X13: 관계사/계열사의 유무, X14: 배우자 직업, X15: 주택

평수, X16: 주택소유자, X17:주거형태, X18: 주택소유기간, X19: 주택거주기간, X22: 당좌거래 은행 수, X23: 은행 외 거래 금융기관개수, X25: 연소득액, X26: 연채산세, X28:부가세)을 후보입력변수로 하여 의사결정나무분석을 수행하였다. 그 결과 아래 <그림3>와 같은 구조를 가진 의사결정나무 구조와 그에 따른 기업의 건전과 부실을 결정짓는 의사결정규칙(decision rule)들을 도출 하여 <그림 3>내의 표와 같이 정리하였다. 소기업 수와 신뢰도는 각각의 의사결정 규칙에 의해 분류된 소기업의 수와 분류 신뢰도를 표현한 것이다.

<그림3> Pruning전 의사결정나무구조 및 가지별 데이터 수와 신뢰도



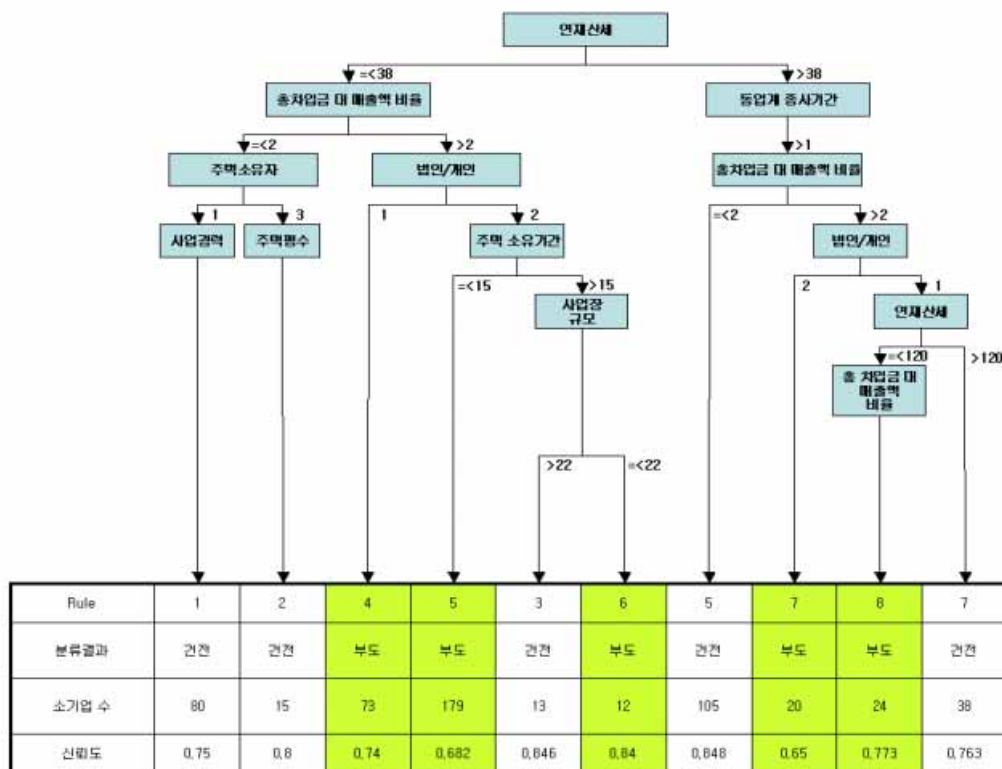
분석결과를 살펴보면 소기업 신용에 가장 큰 영향을 주는 변수는 연채산세를 얼마를 내느냐이며, 총차입금 대 매출액 비율, 경영자 동업계 종사기간, 주택소유자,법인/개인의 여부, 경영자의 사업경력, 사업장규모의 변수들이 부도기업과 건전기업을 판별하는 주요 변수로 파악되었다.

2. 의사결정나무의 구축 - Pruning후

의사결정나무 분석에 의해 도출된 의사결정규칙들은 609개의 분석용 자료(training data)에만 의존하여 구축된 규칙들이므로, 새로운 자료의 예측에 있어 그 안정성이 다소 떨어질 가능성이 있다. 따라서 마지막 가지에 포함된 데이터의 수가 너무 적거나 신뢰도가 낮아 오분류율이 높은 가지에 대해서 Pruning을 통해 의사결정규칙의 일반화를 꾀할 필요가 있다.

따라서 본 연구에서는 마지막 가지에 포함된 데이터의 수가 10개 미만이거나 10개 이상인 경우라도 신뢰도가 낮아 바르게 건전이나 부도로 바르게 분류된 데이터의 수가 10개 미만인 가지의 경우는 의사결정규칙에서 제외시키는 Pruning을 실시하였다. 그 결과 부도기업을 판별하는 Rule중 Rule 1, 2, 3이 Pruning되었고 건전기업을 판별해내는 Rule중에서는 Rule 4, 6이 Pruning되었다. Pruning 후의 의사결정나무 구조는 아래의 <그림4>와 같이 표현할 수 있다.

<그림4> Pruning후 의사결정나무구조 및 가지별 데이터 수와 신뢰도



3. 건전/부도기업을 판별하는 Rule의 생성

의사결정나무의 마지막 가지는 “If~, then~”방식의 Ruleset으로 다음의 <표 11>와 같이 정리 해 볼 수 있다. 여기서 도출된 모든 Rule들이 절대적으로 부실기업과 건전기업을 판별하는데 기준이 된다고는 볼 수 없으나, Pruning과 검증용 데이터에 대한 테스트를 통해 어느 정도 일반성을 획득한 Rule이라 볼 수 있다.

도출된 의사결정규칙들 중 몇 가지 개별 규칙에 대해 살펴보면 다음과 같다. 우선, 건전기업의 분류 Rule1은 ‘만약 연 재산세로 38만원 이하를 내고 총차입금 대 매출액 비율이 15%미만으로 비교적 낮으며, 주택을 소유하고 있는 사람이 본인이어서 주택을 담보로 삼을 수 있고, 대표자로서 동업종에서 종사한 기간이 5년 이상으로 경력이 긴 편에 속한다면, 그 대표자가 이끄는 회사는 건전 기업일 가능성이 높다’와 같이 해석할 수 있다. 반면, 부도기업의 분류 Rule 5를 살펴보면, ‘만약 연재산세로 38만원 이하를 내고 총차입금 대 매출액비율이 15% 이상으로 높고, 그 기업이 개인 기업이고, 기업대표가 주택을 소유한 기간이 15년 이하라면, 그 기업은 부도기업 일 가능성이 높을 것이다’와 같이 해석할 수 있다.

<표11> Ruleset의 정리

건전기업을 판단하는 규칙	
Rule #1 (건전): if 연재산세 =< 38 and 총차입금 대 매출액비율 =< 1 and 주택소유자 == 1 and 사업경력 > 3 then -> 1 (80.0, 0.75)	만약 연재산세로 38만원 이하를 내고 총차입금 대 매출액비율이 15%미만이고, 주택을 소유하고 있는 사람이 대표자 본인이고, 대표자로서 사업경력이 7년 이상 이라면, 그 기업은 건전기업 일 것이다. (포함 소기업수: 80개, 신뢰도:75%)
Rule #2 (건전): if 연재산세 =< 38 and 총차입금 대 매출액비율 =< 2 and 주택소유자 == 3 and 주택평수 > 32 then -> 1 (15.0, 0.8)	만약 연재산세로 38만원 이하를 내고 총차입금 대 매출액비율이 15%미만이고, 주택을 소유하고 있는 사람이 대표자의 배우자이고, 소유하고 있는 주택의 평수가 32평 이상이라면 기업은 건전기업 일 것이다. (포함 소기업수: 15개, 신뢰도:80%)

<p>Rule #3 (건전):</p> <p>if 연재산세 =< 38 and 총차입금 대 매출액비율 > 1 and 법인/개인 == 2 and 주택의소유기간 > 15 and 사업장 규모 > 22 then -> 1 (13.0, 0.846)</p>	<p>만약 연재산세로 38만원이하를 내고 총차입금 대 매출액비율이 5%미만이고, 개인기업이고, 주택을 소유한 기간이 15년 이상이고, 작업장의 평수가 22평 이상이라면, 그 기업은 건전기업 일 것이다.</p> <p>(포함 소기업수: 13개, 신뢰도:84.6%)</p>
<p>Rule #5 (건전):</p> <p>if 연재산세 > 38 and 동업계 종사년수=< 1 and 총차입금 대 매출액비율 =< 2 then -> 1 (105.0, 0.848)</p>	<p>만약 연재산세로 38만원 이상을 내고 대표자가 동업계 종사년수가 3년 이하이고, 총차입금 대 매출액비율이 15% 미만 이라면, 그 기업은 건전기업 일 것이다.</p> <p>(포함 소기업수: 105개, 신뢰도:84.8%)</p>
<p>Rule #7 (건전):</p> <p>if 연재산세 > 120 and 동업계 종사년수> 1 and 총차입금 대 매출액비율 > 2 and 법인/개인 == 1 then -> 1 (38.0, 0.763)</p>	<p>만약 연재산세로 120만원이상을 내고 동업계 종사년수가 3년 이상이고, 총차입금 대 매출액비율이 5% 이상 이고, 그 기업이 법인기업 이라면, 그 기업은 건전기업 일 것이다.</p> <p>(포함 소기업수: 38개, 신뢰도:78.3%)</p>
<p>부도기업을 판별해 내는 규칙들</p>	
<p>Rule #4 (부도):</p> <p>if 연재산세 =< 38 and 총차입금 대 매출액비율 > 2 and 법인/개인 == 2 then -> 2 (73.0, 0.74)</p>	<p>만약 연재산세로 38만원 이하를 내고 총차입금 대 매출액비율이 15% 이상 이고, 그 기업이 개인 기업이라면, 그 기업은 부도기업 일 것이다.</p> <p>(포함 소기업수: 73개, 신뢰도:74%)</p>
<p>Rule #5 (부도):</p> <p>if 연재산세 =< 38 and 총차입금 대 매출액비율 > 2 and 법인/개인 == 2 and 주택의소유기간 =< 15 then -> 2 (179.0, 0.682)</p>	<p>만약 연재산세로 38만원 이하를 내고 총차입금 대 매출액비율이 15% 이상 이고, 그 기업이 개인 기업이고, 주택을 소유한 기간이 15년 이하라면, 그 기업은 부도기업 일 것이다.</p> <p>(포함 소기업수: 179개, 신뢰도:68.2%)</p>

<p>Rule #6 (부도):</p> <p>if 연채산세 \leq 38 and 총차입금 대 매출액비율 $>$ 2 and 법인/개인 == 2 and 주택의소유기간 $>$ 15 and 사업장 규모 \leq 22 then \rightarrow 2 (10.0, 0.6)</p>	<p>만약 연채산세로 38만원 이하를 내고 총차입금 대 매출액비율이 15% 이상 이고, 그 기업이 개인일 경우, 주택의 소유기간이 15년이상 이긴 하나 사업장 규모가 22평이하의 소규모라면 그 기업은 부도기업 일 것이다.</p> <p>(포함 소기업수: 10개, 신뢰도:60%)</p>
<p>Rule #7 (부도):</p> <p>if 연채산세 $>$ 38 and 동업계 종사년수 $>$ 1 and 총차입금 대 매출액비율 $>$ 2 and 법인/개인 == 2 then \rightarrow 2 (20.0, 0.65)</p>	<p>만약 연채산세로 38만원 이상을 내고 동업계 종사년수가 3년 이상이긴 하나, 총차입금 대 매출액비율이 15% 이상 이고, 그 기업이 개인기업 이라면 그 기업은 부도기업 일 것이다.</p> <p>(포함 소기업수: 20개, 신뢰도:65%)</p>
<p>Rule #8 (부도):</p> <p>if 연채산세 $>$ 38 and 연채산세 \leq 120 and 동업계 종사년수 $>$ 1 and 총차입금 대 매출액비율 $>$ 4 and 법인/개인 == 1 then \rightarrow 2 (24.0, 0.773)</p>	<p>만약 연채산세로 38만원 이상 120만원이하를 내고 동업계 종사년수가 3년 이상 이고, 총차입금 대 매출액비율이 40% 이상 이고, 그 기업이 개인기업 이라면 그 기업은 부도기업 일 것이다.</p> <p>(포함 소기업수: 24개, 신뢰도:77.3%)</p>

4. 예측 정확도

아래의 표들은 의사결정나무 기법의 예측 퍼포먼스를 나타내는 소기업 부실/건전 판별 적중률이다. 우선 <표12>는 훈련용 데이터의 부실/ 건전의 적중률로서 의사결정나무 기법 전체의 분류 정확도는 79.15%, 실제 건전기업을 건전기업으로 예측하는 특이도는 81.17%, 실제 부실기업을 부실기업이라고 예측하는 민감도는 76.84%임을 알 수 있다. 나머지는 검증용 데이터 1,2,3,4에 대한 부실/ 건전의 적중률을 나타내는 것으로 대부분 특이도가 민감도에 비해 상대적으로 예측력이 떨어지지만 전반적인 분류 정확도는 약 73% ~78% 까지 대체로 일정한 정확도를 보이고 있으므로 의사결정나무 기법을 이용해서 구축한 소기업 신용평가 모형이 어느 정도 일반성을 확보했다 판단할

수 있다.

<표12> 훈련용 및 검증용 데이터의 부도/건전 적중률

훈련용 데이터		예측값		합계	예측력	
		건전	부도			
실제값	건전	263	61	324	특이도	81.17%
	부도	66	219	285	민감도	76.84%
합계		329	280	609	정확도	79.15%
검증용 데이터1		예측값		합계	예측력	
		건전	부도			
실제값	건전	27	14	41	특이도	65.85%
	부도	10	39	49	민감도	79.59%
합계		37	53	90	정확도	73.33%
검증용 데이터2		예측값		합계	예측력	
		건전	부도			
실제값	건전	31	14	45	특이도	68.89%
	부도	8	37	45	민감도	82.22%
합계		39	51	90	정확도	75.56%
검증용 데이터3		예측값		합계	예측력	
		건전	부도			
실제값	건전	31	14	45	특이도	68.89%
	부도	7	38	45	민감도	84.44%
합계		38	52	90	정확도	74.44%
검증용 데이터4		예측값		합계	예측력	
		건전	부도			
실제값	건전	32	13	45	특이도	71.11%
	부도	7	38	45	민감도	84.44%
합계		39	51	90	정확도	77.78%

VI. 결론

본 연구에서는 최근 관심은 증대되고 있으나 축적된 심사 경험의 부족과 고객데이터의 한계 때문에 구축에 어려움을 겪고 있는 소기업 신용평가 모형을 구축하였다. 이를 위해 신용평가 전문가들의 토의를 거쳐 완성된 설문을 이용하여 소기업 데이터를 직접 수집하였고, 질적 정보의 분석에 용이한 의사결정나무를 활용하였다. 즉, t-검정, Mann-Whitney-Wilcoxon검정, χ^2 -검정과 같은 통계분석을 실시하여 건전/부실기업에 대한 각 변수들의 유의성을 검정하여 입력변수를 선정하고 의사결정나무 기법을 통해 부도와 건전을 판별하는 규칙들을 추출하는 방식으로 모형을 구축할 수 있었다.

여기서 구축된 모형은 활용된 데이터들이 기업별 특성이 강한 소기업을 대상으로 했음에도 불구하고 대체로 훈련용 데이터와 검증용 데이터 모두 73~79%의 안정된 예측 정확도를 보이고 있다. 이는 앞으로 소기업에 대한 좀 더 체계적인 데이터 축적이 이루어진다면, 데이터를 기반으로 한 소기업 신용평가 모형 구축이 질적으로 향상할 수 있다는 가능성을 제시하고 있는 것이다.

따라서 본 연구는 이제까지 시도된 적이 거의 없었던 소기업 신용평가과정을 실제 C은행의 소기업 고객들에 대한 데이터를 직접 수집하고 의사결정나무 기법을 활용하여 구조화함으로써 데이터를 기반으로 한 소기업 신용평가 모델 개발의 구축 가능성을 제시했다는 점에 의의가 있다.

그러나 본 연구는 다음과 같은 몇 가지 점에서 한계가 있다.

첫째, 데이터 수집상의 한계가 있다는 점이다. 국내에서는 현재까지 소기업의 재무적, 비재무적 데이터는 금융기관들 역시 체계적으로 정리하지 않은 상태이거나 이를 위한 준비를 하고 있는 상태이다. 본 연구에서 소기업의 데이터로 C은행의 소기업 고객들을 대상으로 실시한 survey결과를 이용하였지만 수집된 샘플기업의 반 이상이 중요한 변수들의 결측값을 가지고 있어 모든 데이터를 충분히 활용하는 데 한계가 있었다.

둘째, 업종별 특성을 반영한 정교한 신용평가 모형을 구축하지 않았다는 점이다. 이는 업종별로 충분한 데이터가 확보되지 않은 데서 기인한 것으로, 앞으로 소기업 신용평가가 활성화되어 충분한 데이터가 확보 된다면, 세분화된 분석을 통해 업종별 특

성을 반영한 보다 효과적인 신용평가 모형의 구축이 가능할 것이다.

셋째, 가지치기 방법과 정지규칙의 모든 경우의 수에 대해 실험할 수 없었다는 점이다. 의사결정 모형에는 모형의 일반성과 정확성을 높이기 위한 여러 가지치기방법(pruning)과 정지 규칙(stopping point)이 있다. 그러나 본 연구에서는 가지치기 방법과 정지 규칙의 모든 경우의 수에 대해 실험하는 것에 한계가 있어 연구자가 임의로 75% severity 가지치기와 한 Parent 당 최소 10개 이상의 데이터가 포함되도록 하는 정지 규칙으로 고정시킨 후 실험을 실시 하였다. 따라서 향후에는 가지치기나 정지규칙을 다양하게 변경하여 실험을 실시하고 비교를 통해 최적의 조건을 찾아 모형을 구축하거나 전문가의 지식을 반영한 휴리스틱스(Heuristics)를 가지치기의 기준으로 적용해 보는 것도 좋은 연구과제가 될 수 있을 것이다.

참고문헌

국내 참고문헌

- [1] 기현희(1994) <중소기업의 부도예측모형에 관한 실증적 연구>, 서울여자대학교.
- [2] 김영남(2002) <데이터마이닝의 의사결정나무분석을 이용한 사례분석, 이화여자대학교, 통계대학원, 통계
- [3] 김용기(1983) <신용조사>, 무역경영, pp.155-160.
- [4] 김종재, <중소기업경영론> 박영사, p.9.
- [5] 김지현(2001) <다중나무를 이용한 사례분석>, 이화여자대학교, 통계대학원, 통계학전공
- [6] 김충섭(1997) <중소기업 신용평가의 질적 요인에 관한 실증적 연구>, 고려대학교 통계학과.
- [7] 김형수(1999) <신용평가의 기능과 행정적 활용>, 한국신용평가주 선임연구원.
- [8] 김희열(1993) <체계경영분석>, 박영사.
- [9] 김희진(2001) <E-miner 와 Clementine을 이용한 의사결정나무 사례분석>, 이화여자대학교, 통계대학원
- [10] 민경선(1989) <기업신용평가에 있어서 비재무적 용인의 중요성에 관한 연구>, 고려대학교 경영대학원
- [11] 박상범(1998) <중소기업론>, 삼영사, pp.24-25.
- [12] 박재린(1999) <최신 중소기업경영론>, 무역경영사.
- [13] 박중원(2001) <우리나라 신용평가제도의 문제점과 개선방안에 대한 연구>.
- [14] 박중원(2001) <우리나라 신용평가제도의 장단점과 개선방안에 대한 연구>, 동아대학교 경영대학원
- [15] 손상호, 김동환(1999) <중소기업 신용평가제도 개선방안>, pp.19- 25.
- [16] 손정식 외2인(1992) <금융기관경영론>, 박영사, pp.245-246.
- [17] 손정현(1998) <우리나라 신용평점 모형개선책 연구>, 연세대학교 경영대학원, 석사학위논문, p.75.
- [18] 송문섭, 박창순(1989) <비모수통계학개론>, 자유아카데미, pp.50-54.
- [19] 신동령(1999) <신용평가의 실제와 이론>, 다산출판사.
- [20] 원태연, 정성원 공저 <한국 SPSS10K 통계조사분석>, SPSS 아카데미.

- [21] 이윤신(2000) <중소기업 신용평가제도에 관한 실증적 연구>, 서강대학교 경제대학원 금융경제전공.
- [22] 이진우(2000) <중소기업 신용평가 시 비재무정보의 유용성에 관한 연구>, 한양대학교.
- [23] 임건길(1991) <중소기업 신용평가에 있어 비재무정보의 유용성에 관한 실증적 연구>, 고려대학교.
- [24] 윤태순(1999.8) <국내은행의 신용평가 모형 개선과 효과적인 Credit Rating System 구축방안>, 월간조흥경제
- [25] 전명식(1996) <수리통계학>, 자유아카데미, pp.237-239.
- [26] 정충영, 최이규(2000) <SPSS WIN을 이용한 통계분석>, 무역경영사, p.282.
- [27] 조옥래(1998) <계량심리학의 방법론을 이용한 체계적인 전문가 지식구조분석 방법: 비재무항목을 활용한 중소기업 신용평가 전문가시스템 규칙개발에 적용>, 경희대학교 대학원
- [28] 조옥래(1998) <중소기업도산예측과정에서 비재무지표항목의 유용성검증과 기업 분석전문가 시스템 개발에 관한 연구>, 경희대학교, 경영대학원, 회계학과
- [29] 조희영, 박상범 공저(1997) <중소기업경영론>, 삼영사, p.24 .
- [30] 최성운(2002) <인공신경망을 이용한 소기업 신용평가 모형에 관한 연구>.
- [31] 최용식(1999) <IMF이후의 기업 금융론>, 전영사.
- [32] 한국금융연구원(1999) <중소기업 신용평가제도 개선방안>.
- [33] 한인구,신경식(1998),"지능형 중기업 신용평가시스템의 개발 및 활용: 보람은행의 사례를 중심으로"
- [34] 한희영(1985) <경영학총론>, 다산출판사, p.219.
- [35] 허명희, <SAS 범주형 자료분석>, 자유아카데미, pp.3-5.
- [36] 현점휴(1998) <대출의사결정에 있어 회계정보의 유용성에 관한 연구>.

외국 참고문헌

- [1] Altman, E.I. (1968) <Financial ratios>, Discriminant analysis and the prediction of corporate bankruptcy, The journal of Finance, September, p.589-609.
- [2] Altman, E. I., Haldeman, R. G. and Narayanan, P. (1997) <ZETA Analysis>, A New Model to Identify Bankruptcy Risk of Corporation, Journal of Banking and Finance, June, p.29-54.
- [3] Breiman, Freidman, Olshen, Stone (1984) <Classification and Decision Trees>, Wadsworth.
- [4] Boritz, J. and Kennedy, D.(1995) <Effectiveness of neural networks types for prediction of business failure, Expert Systems with Applications 9, pp503-512
- [5] Bell, T.(1997) <Neural nets or the logit model: A comparison of each model's ability to predict commercial bank failures," Intelligent Systems in Accounting Finance and Management, Vol.6, pp249-264
- [6] Cadden, D. T.(1991) <Neural Network and Mathematics of Chaos> -An Investigation of These Methodologies as Accurate Predictors of Corporate Bankruptcy, Proceedings of the First International Conference on Artificial Intelligence Application on Wall Street, p.52-57.
- [7] Clifford. M. Baumbach(1985) <How to organize and operate a small business> Prentice-Hall, inc., p.3.
- [8] Deakin, E. B. (1972) <A Discriminant Analysis of Predictor of Business Failure>, Journal of Accounting Research, 10.
- [9] Edward I. Altman(1968), <Financial Rations, Discriminant Analysis and the Prediction of Corporate bankruptcy>, The Journal of Finance, Vol.23, No.4, September, p.12-27.
- [10] Edward B. Deakin,(1972) <A Discriminant Analysis of Predictors Business Failure>, The Journal of Accounting Research, p.167-169.
- [11] Gim, G., Whalen, T(1996) <Second Order Logical System for Risk Classification in Newly Developed Country>, International Journal of Uncertainty, Fuzziness and Knowledge Based System, Vol.4, No.5, p.421-430.
- [12] Gentry, J., Newbold, p. and Whitford, D., <Classifying bankruptcy firms with fund flow components," Journal of Accounting Research, spring, 1985, pp123-133

- [13] J. Ross Quinlan (1993) <C4.5: Programs for Machine Learning> The Morgan Kaufmann Series in Machine Learning, Pat Langley, Series Editor.
- [14] J. R. Quinlan (1996) <Improved Use of Continuous Attributes in C4.5>- Journal of Artificial Intelligence Research 4 p.77-90.
- [15] J. O. Horrigan,(1965) <Some empirical bases of financial ration analysis>
- [16] J. Argenti,(1977) <Company Failure: long-range prediction not enough>, Accountancy, August, p.52.
- [17] J. Ohlson,(1980) <Financial rations and the probabilistic prediction of bankruptcy>, Journal of Accounting Research,
- [18] Kass, G(1980) <An Exploratory Technique for Investigating Large Quantities of Categorical Data>, Applied Statistics, Vol. 29, p.119-127.
- [19] Ken, S. C, and Linda, C. C(1998) <Marjet Structure and Discrimination: The Case of Small Businesses>, Journal of Money, Credit and Banking, Vol.30, No. 4, p.771-792.
- [20] Kim, H, Koehler, G. J.(1995) <Theory and Practice of Decision Tree Induction>, Omega, Vol. 23, No. 6, p.637-652.
- [21] Michael J. A. Berry and Gordon S. Linoff,(1997) <Data Mining Techniques for marketing, Salesm and Customer Support>, p.243-285.
- [22] Melody Y. Kiang(2002) <A Comparative Assessment of Classification Methods>, Decision Support Systems.
- [23] Proving and Integrating Educational Resources for Faculty Teaching Artificial Intelligence: held at Temple University in Philadelphia, (1994) <Building Classification Models: ID3 and C4.5>.
- [24] Quinlan, J. R. (1987) <Simplifying decision trees>- international Journal of Man-Machine Studies, p. 27, p.221-234.
- [25] Shin, K. S., Shin, T. and Han, I. (1998a) <Corporate Credit Rating system Using Bankruptcy Probability Matrix>, Proceedings of the IV International Meeting on Artificial Intelligence and Emerging Technologies in Accounting, Finance a Taxation on Huelave, Spain.
- [26] Shin, K. S., Shin, T, And Han, I. (1998b) <Neuro-genetic Approach for Bankruptcy Prediction>, A Comparison to Back-propagation Algorithms, International Conference of Korea Management Information System Society,

p.585-597.

[27] Seiden, M. H.(1962) <Trade Credit: A Quantitative and Qualitative Analysis>, 42nd Annual Report, National Bureau of Economic Research.

[28] Zmijeswski, M. E.(1984) <Methodological Issues Related to the Estimation of Financial distress Prediction Models>, Journal of Accounting Research, No,22, pp59-82

Abstract

A Decision Tree Approach

For Small Enterprise Credit Evaluation Model

Lee, Eui Su
Dept. of Business Administration
The Graduate School of Ewha Womans University

As the recent rapid changes of circumstances progress in the Korean financial market, the competition between Korean financial institutions became severer. This makes for Korean financial institutions not to manage their fund to Chaebols under the help of government any more. To counter the challenges posed by the changing environment, domestic institutions need to rationalize their management style. So some of them try to expand range of customers to small- medium sized enterprises (SMEs) for the effects of lowering risks and raising return.

But for the case of SMEs, it is hard to gather the good quality of financial information and the cost of credit evaluation is high when it comes to the situation that there are lots of demands for evaluation with the small loan.

Therefore, it is necessary to develop objective and rational credit evaluation model for evaluating SMEs more effectively and efficiently. In spite of this demand, both researchers and financial institutions were having difficulties to develop due to the lack of data and enough experiences.

This thesis takes this lost opportunity as a starting point. For the purpose of my study, a SME is defined as a company with total assets of less than 1 billion won which is current or past customer of C Bank.

For the research, I look into the theoretical background of general credit

analysis model and its application to SMEs. First, I gather data of 789 SMEs including healthy and unhealthy companies from the survey data of C bank's SMEs done during last Dec. 2001 under.

Then I select input variables for Decision Tree using t-test, Mann-Whitney-Wilcoxon test, χ^2 -test and build credit evaluation model for SMEs utilizing Decision Tree C 5.0 algorithm.

The result of a new model demonstrated accuracies between 73~79% from training data and verifying data respectively. This result is meaningful that it shows the possibility to improve the credit evaluation process for SMEs more effectively.

I believe that the utilization of the SME credit evaluation model will give banks competitive business tools and contribute to stimulate Korean economy.