

## 전자상거래에서 지식탐사기법의 활용에 관한 연구\*

성태경\*\* · 주석진\*\*\* · 김중환\*\*\*\* · 홍준석\*\*\*\*\*

### < 목 차 >

I. 서론	4.2 연구변수의 선정
II. 문헌연구	4.3 표본의 분할
III. 모형추정용 기법들의 특성	4.4 도산예측모형
3.1 판별분석	V. 연구결과의 의미
3.2 의사결정나무	VI. 요약 및 결론
3.3 신경망 기법	참고문헌
IV. 연구방법론	Abstract
4.1 표본의 선정	

## I. 서론

기업의 도산은 기업인이나 주주에게는 재산의 상실을, 노동자에게는 실업을, 일반 국민들에게는 취업 기회의 감소 등의 경제적 손실을 유발할 뿐 만 아니라 이들 기업들을 상대로 다양한 거래를 유지하던 금융기관들의 정상적인 운영에 장애로 작용한다. 이는 최근 합병이나 흡수의 형태로 사라져간 많은 금융기관들이 부실했던 주요 원인 중의 하나가 도산할 기업에 대한 무분별한 대출과 대출금 회수를 저조였다는 점에서 증명된다. 따라서 효과적인 위험관리를 위해 정확한 기업 도산 예측모형을 준비하고 활용하고자 하는 연구들이 현업과 학계를 중심으로 꾸준히 수행되어 왔다. 그러나 지금까지의 연구는 전통적인 기업에 집중되어왔고, 20세기말부터 등장한 전자상거래 혹은 온라인 기업의 도산에 대한 연구는 전무한 상태이다.

전 세계는 지금 인터넷상의 가상사회를 매개로 한 새로운 거래방식인 전자상거래의 폭발적 증가를 경험하고 있다. 대표적으로 Amazon.com은 자본금 250억불 규모의 기업으로 성장하여, 우리나라 주식시장 가치의 50% 규모에 육박하였으며, GE는 사이버 입찰을 통하여 연간 50억불 이

\* 본 연구는 2003학년도 경기대학교 학술연구비 (연구소과제) 지원에 의하여 수행되었음

\*\* 경기대학교 경영학부 경영정보학전공 교수, tksung@kyonggi.ac.kr

\*\*\* 경기대학교 경영학부 경영정보학전공 교수, sju@kyonggi.ac.kr

\*\*\*\* 경기대학교 경영학부 경영정보학전공 교수, jhkim@kyonggi.ac.kr

\*\*\*\*\* 경기대학교 e-비즈니스연구소 연구원, junehong@kyonggi.ac.kr

상을 구매함으로써 10% 이상의 구매비용 절감과 40%의 구매시간의 단축을 실현하였다고 보고 되고 있다. 삼성경제연구소의 자료에 따르면, 세계 전자상거래 시장규모가 매년 180% 성장하고 있으며, 우리나라는 매년 200% 이상씩 성장하여 2005년에는 2조 6000억 원대에 이를 것이라고 예측하고 있다. 따라서 전자상거래 혹은 온라인 기업의 부도에 대한 연구의 중요성과 시급성이 강조되고 있는 현실이다.

기업 도산예측에 관한 연구는 Altman (1968)의 Z score 모형을 기점으로 다변량 판별분석 (Multi-variate Discriminant Analysis), 로짓회귀분석 (Logistic Regression Analysis) 등의 통계학적 기법, 신경망 (Neural Network), 의사결정나무 (Decision Tree) 등의 인공지능 기법 등 지식탐사기법 혹은 데이터마이닝을 이용한 다양한 기법들이 도산 예측모형 개발에 사용되었다 (Berry and Linoff, 1997). 그러나 대부분이 전통적인 기업의 도산 자료를 토대로 만들어진 모형으로, 이들은 현재 급속한 성장을 나타내고 경제를 주도하는 전자상거래 혹은 온라인 기업에 적용할 경우 모형의 적합성 및 예측의 신뢰도에 문제가 있으리라 예견된다.

본 연구의 목적은 전통적인 기업의 도산 예측모형과 전자상거래 혹은 온라인 기업의 도산 예측모형을 개발하여, 기업도산 예측모형의 유동성을 살펴보고 모형의 변화에 영향을 미치는 다양한 요소들을 비교 분석함으로써 기업의 성격에 따른 새로운 모형의 필요성을 제시하는데 있다. 이를 위하여 모형의 해석이 가능한 의사결정나무와 다변량 판별분석 기법을 모형의 추정을 위해 선택하였으며 상호 비교 평가하였다. 논문의 구성은 다음과 같다. 2장에서는 기업도산 예측에 관한 과거 연구 내용을 문헌연구를 통해 조사하였으며, 3장에서는 도산 예측모형 추정에 사용될 수 있는 대표적인 기법들의 특성을 서술하였다. 4장에서는 자료 수집 및 변수 선정 과정 등을 논하고 판별분석과 의사결정나무를 통해 각각 추정된 모형을 비교 평가하였다. 5장에서는 모형의 추정을 통해 지목된 변수들이 의미하는 바를 해석하고, 현재 국내 금융기관에서 기업의 신용평가를 위해 사용하는 변수들과의 차이를 조사하였다.

## II. 문헌연구

기업의 도산 예측에는 재무비율 분석이 널리 이용되어 왔는데, Altman에 의하면 최초의 연구는 1935년 대공황 시기에 Smith와 Worker에 의해 이루어 졌으며, 1942년 Mervin은 도산한 기업의 재무비율이 정상기업에 비해 현저히 다르다는 것을 밝혔다 (Altman, 1993). Neter (1967)는 자신의 논문에서 재무비율을 이용하여 기업도산을 예측하기 위하여 다변량 분석 (Multi-variate Analysis) 기법을 Beaver (1966)는 단일변량분석 (Uni-variate Analysis) 기법을 제시하였다.

그러나 실질적으로 활발한 연구가 시작된 것은 60년대 후반 Altman (1968)이 Z score 모형을 발표한 이후이다. Z score모형은 다중 판별분석 (Multiple Discriminant Analysis)에서 얻어진 5개의 변수로 구성되었는데 90% 이상의 높은 예측력을 제공하였으며, 이러한 결과를 배경으로 당대 가장 보편적인 도산 예측모형 추정을 위한 접근법으로 인식되었다. 그러나 Altman의 Z

score 모델을 포함한 대부분의 연구는 주로 소규모 사업체를 대상으로 이루어졌기 때문에 이를 일반 사업체에 확대 적용하기에는 문제가 많았다. Altman, Haldeman, Narayanan (1977) 등은 Z score 모델을 발전시켜 특정 규모에 관계없이 일반 사업체에 적용할 수 있는 'ZETA'라는 모델을 발표했으며 Platt (1990, 1991), Weiss (1981), Zavgren (1988) 등은 다중 판별분석 기법이 제공하는 도산 예측율의 문제점을 제기하기도 했다.

Wilcox (1991)는 도산예측에 Gambler's Ruin 모델을 적용하여 Beaver (1966)의 연구결과를 보강하고 자신이 제시한 모델의 이론적 타당성을 주장하였으나 크게 주목받지는 못했다. Scott (1981)는 기존에 발표된 다양한 모델들을 이론적 적합성과 예측율의 정확도 관점에서 비교 연구하였다. 그는 Altman, Haldeman, and Narayana (1977) 등이 제시한 다중 판별분석 기법에 근거한 'ZETA' 모델이 수익과 부채에 대한 변수를 지니고 있고, 회계와 주식 관련 자료를 포함하며, 간결하면서도 탁월한 판별력을 제공하기 때문에 가장 신뢰할 수 있는 모델이라고 결론 내렸다. 이밖에도 기업의 도산예측 모형을 위해 Olson (1980)은 로짓회귀분석 (Logit Regression Analysis)을, Zmijewski (1984)는 프라빗분석 (Probit Analysis)을, 그리고 West (1985)는 요인 분석 (Factor Analysis)을 활용하는 등 다양한 통계기법이 사용되었다.

80년대 중반이후에는 신경망 (Neural Nets), 의사결정나무 (Decision Trees) 기법 등 인공지능을 기반으로 한 지식탐사기법 혹은 데이터마이닝 기법들이 기업의 도산 예측모형에 활발히 적용되고 있는데 이들은 일반적인 통계학적 가정을 필요로 하지 않는다는 점이 특징이다 (Berry and Linoff, 1997). 이 중에서도 신경망 기법은 도산예측에 가장 널리 사용된 기법인데 대부분의 연구들이 후진전파 (Backpropagation) 기법을 토대로 모형을 추정하고 의사결정나무, 판별분석, 유전 (Genetics) 알고리즘 등 타 기법과의 성능 (주로 예측력) 비교에 초점을 맞추고 있다 (Altman et. al., 1994; Chung and Hong, 1995; Dutta and Shekhar, 1988; Klimasauskas, 1993; Lee, 1993; Lee et. al, 1995; Odom and Sharda, 1990; Raghupathi et. al., 1991; Rahimain et al., 1993; Sung and Han; 1998; Tam and Chi, 1991; Wilson and Sharda, 1994). 최근에는 1986년에 발표된 Quinlan (1986)의 ID3와 이를 계승 발전시킨 C4.5 알고리즘을 주로 적용하는 의사결정나무 기법도 도산 예측모형에 사용되고 있으며 (Quinlan, 1996a and 1996b), 유전적 원리와 진화론적 이론에 근거한 유전 알고리즘도 간헐적으로 선택되고 있다 (Michalewicz, 1984). 특히 유전 알고리즘의 경우 단독으로 보다는 신경망 기법을 적용하기 전 단계에서 변수의 선택을 위해 사용하는 사례를 접할 수 있다 (이건창, 1995; Goldberg, 1989; Lee et. al, 1995; Shin et al., 1996).

앞서 조사한 연구의 대부분이 정상적인 경제 환경을 가정하여 예측모형을 제시하고 있는 반면, 조직이론이나 의사결정 등의 관리 (Behavioral) 분야에서는 경제위기 상황을 함께 고려한 연구들이 종종 발표되고 있다. Smart와 Vertinsky (1977 and 1984) 는 경제위기 상황에서 금융권이 기업의 도산으로 인해 발생하는 회수 불가능 부채로부터 재정상의 손실을 최소화하기 위한 방안으로서의 경영전략을 기업의 신용평가 관점에서 제시하였다. Bryson (1981)은 자신이 제창한 기회공간 (Opportunity Space) 이론에서 정상적인 경제 환경에서 의사결정을 위해 고려하는 제약이나 법칙들은 다른 환경에서는 반드시 수정되어야 한다고 주장하고 있다. Billing, Milburin,

Schaalman (1980) 등은 위기 상황의 특징과 이에 대처하는 의사결정자들의 행동에 대한 연구를 발표했는데, 불확실성이 증대되고 빠른 의사결정을 요하며 잘못된 의사결정에 따른 비용이 엄청날 수밖에 없는 경제위기 환경 하에서 의사결정자들은 올바른 선택에 대한 중압감으로 인해 판단착오를 일으키기 쉬우며 정상적일 때 사용하던 것과는 다른 새로운 의사결정 모형의 설계 및 적용을 통해 이들의 혼란을 최소화시켜야 한다는 것을 강조했다.

1999년 경제개발협력기구 (OECD: Organization for Economic Co-Operation and Development) 는 전자상거래를 새로운 경제·경영의 한 축으로 인식하고, 전자상거래가 전세계 경제 및 사회에 미치는 영향은 일반인들이 생각하는 것보다 훨씬 크다는 보고서를 발표한바 있다. 이제 전자상거래는 자유경제를 활성화하고 무역장벽이라는 장애를 극복할 수 있는 매력적인 대안 경제로 각광을 받고 있다 (Brynjolfsson and Kahin, 2000, Adam, et. al, 1999; Westland and Clark, 1999). 미국 통산성에 따르면 2006년에 미국 고용인구의 절반 이상이 정보기술과 관련된 직종에서 일할 것이라고 예측하였고, Cisco와 The University of Texas at Austin 부설 CREC (Center for Research in Electronic Commerce) 의 조사에 따르면 인터넷 경제는 매출액 기준으로 지난 수년간 년 60% 정도의 증가를 보이고 있다 (2001). 이를 반영하듯 미국은 물론 전세계적으로 전자상거래 분야에 진출하는 기업이 급증하고 있다.

그러나 이렇게 급성장하는 전자상거래업계에서 진출한다고 모두 성공하는 것은 아니다. 수많은 Dot.com 기업들의 흥망성쇠를 관찰할 때, 전자상거래 기업의 도산에 대한 연구가 시급한 시점이라고 할 수 있다. 특히 대부분의 전자상거래기업이 벤처의 형태를 보이고 있어, 은행을 포함한 금융기관이 대출의 위험을 낮추기 위해서도 전통적인 기업과는 상이한 도산예측모형이 필요하다. 벤처산업 육성을 우선 과제로 있는 우리나라에서, 벤처의 성공 가능성은 첨단기술 분야나 인터넷/전자상거래 분야가 높다는 점을 고려하면, 전자상거래기업의 도산을 예측하는 모형 개발에 대한 학문적, 실질적 노력이 필요한 시점이라고 할 수 있다.

### Ⅲ. 모형추정용 기법들의 특성

컴퓨터 처리기술의 급속한 발전과 상대적인 가격하락, 그리고 계정계시스템 (Legacy System) 의 지속적인 보급 등으로 인해 다량의 자료가 전자적으로 수집되고 있으며, 이들 자료를 이용하여 모형을 추정하고 의미 있는 정보를 추출하는 지식탐사 (데이터마이닝) 기법이 다양한 영역에 활발히 적용되고 있다. 그러나 많은 경우가 사례에 적합하지 못한 기법을 선택하여 예측모형을 추정함으로써 모형의 부적합성이나 예측결과와 과잉 또는 과소해석을 유발한다는 지적이 최근 들어 일고 있다 (Berry and Linoff, 1997). 따라서 각 기법의 장단점을 조사한 후 적절한 기법을 선택하는 것이 연구의 효율성 및 신뢰성의 제고를 위해 필수적이라 하겠다. 가장 보편적으로 활용되는 판별분석, 의사결정나무, 신경망을 살펴보기로 한다.

### 3.1 판별분석

오랜 역사와 탄탄한 이론을 배경으로 한 검증된 기법으로 자료의 동시 통합/분석을 통하여 분류에 영향을 미치는 주요 변수들을 선별하고 이들의 영향력을 해석 가능한 수식으로 제공하는 능력을 지니고 있다. 특히 연속형 입력변수의, 목표변수 (분류/종속변수)에의 기여도에 대한 설명력은 타 기법들에 비해 월등하다. 반면, 자료 값들의 정규분포, 각 집단 (도산, 비도산)의 균등 공분산 (Covariance) 등 여러 가지 통계학적 가정을 요구하고 자료에 다수의 범주형 변수가 포함되어 있을 때 이들의 처리과정에서 발생할 수 있는 치우침 (Bias) 등의 한계를 지니고 있다.

### 3.2 의사결정나무

의사결정나무는 모형추정용 (Training) 자료로부터 순환적분할 (Recursive Partitioning) 방식을 이용하여 나무를 구축하는 기법으로 구축 되어진 나무는 속성의 분리기준을 포함하는 내부마디 (Internal Nodes)와 최종분류를 의미하는 잎 (Leaves)으로 구성되며, 분류나 예측에 주로 사용된다. 이 개념은 Hunt, Marin, Stone (1966)에 의해 처음 소개되었으며 80년대 중반 Quinlan (1986)이 이를 수정하고 변형하여 ID3를 발표한 이후 인공지능 및 기계학습 (Machine Learning)의 다양한 분야, 예를 들어 우량주식 분류 (Norris, 1996), 포트폴리오 분석 (Tam and Chi, 1991), 은행도산예측 (Tam and Kiang, 1992) 등에서 폭넓게 사용되고 있다. 이 기법의 최대 강점은 특정 통계적 가정을 전제로 하지 않으며, 분석을 통하여 얻어지는 정보의 패턴이 나무 또는 If-Then 형식의 법칙으로 만들어 지기 때문에 쉽게 이해할 수 있다는 것이다. 또한 학습을 통한 모형 추정에 걸리는 시간이 짧고, 연속형 변수들과 범주형 변수들을 사전 가공작업 없이 바로 사용할 수 있다. 반면에 나무를 구축할 때 연속형 변수의 경우 값을 분리하는 과정에서 치우침의 우려가 많으며, 빈값 (Missing Value)이나 오류값 (Noise)에 대한 처리능력이 부족하다.

### 3.3 신경망 기법

인간 두뇌의 신경세포를 마디 (Node)와 고리 (Link)를 이용하여 모형화하고, 모형추정용 자료로부터 반복적인 학습과정을 거쳐 자료에 내재되어 있는 패턴을 찾아내는 모델링 기법이다. 현재 가장 널리 사용되는 지식탐사 (데이터마이닝) 기법으로 분류, 예측, 군집 (Clustering), 연관 (Association) 등의 다양한 분석에 광범위하게 적용할 수 있으며, 의사결정나무와 마찬가지로 특정 통계적 가정을 전제로 하지 않는다. 그리고 결과의 예측력도 다른 기법들에 비해 우수한 것으로 알려져 있다 (Odom and Sharda, 1990; Rahimain et al., 1993; Tam and Chi, 1991; Wilson and Sharda, 1994). 그러나 이 기법의 가장 큰 약점은 제공하는 결과를 설명할 수 없는 블랙박스 기법 (Black Box Approach) 라는 것이다. 또한 학습시간이 길며, 입력되는 자료의 값을 0~1 범위 내에서 요구함으로써 인해 사전 자료 가공작업을 필요로 하고, 결과의 값이 입력되는 자료의 순

서에 따라 변한다. 따라서 신경망 기법은 주로 모형의 설명력보다는 결과의 정확도를 요구하는 상황에 적절한 기법이라 하겠다.

## IV. 연구방법론

### 4.1 표본의 선정

앞서 서술한 바와 같이 본 연구의 목적은 전통적인 기업의 도산 예측모형과 전자상거래 혹은 온라인 기업의 도산 예측모형을 개발하여, 기업도산 예측모형의 유동성을 살펴보고 모형의 변화에 영향을 미치는 다양한 요소들을 비교 분석함으로써 기업의 성격에 따른 새로운 모형의 필요성을 제시하는데 있다. 전통적 기업과 전자상거래 (온라인) 기업을 구분하는 일은 기술적으로 상당히 어려운데, 이는 100% 전통적 (off-line) 기업과 100% 전자상거래 (온라인) 기업이 거의 없으며, 온라인과 오프라인을 동시에 추구하는 기업이 대다수이기 때문이다. 따라서 본 연구에서는 매출액의 80%가 오프라인에서 발생하면 전통적 기업으로, 그리고 매출액의 80%가 온라인으로 발생하면 전자상거래기업으로 분류하였다.

자료는 상장회사협의회에서 발행하는 기업 재무제표를 활용하였으며, 도산 기업과 비도산 기업 관련 자료는 한국증권거래소의 데이터베이스를 통하여 조회하였다. 기간은 2002년 1월 1일부터 2004년 12월 31일까지로 하여 자료를 수집하였고, 전자상거래 (온라인) 기업이 전통적 기업보다 비교적 규모가 작다는 점을 고려하여, 채택된 기업의 선정 기준은 제조업체 중 총자산 규모가 50억 이상 500억 이하이고, 종업원 수가 50명 이상 500명 이하로 정했는데 이는 자료의 동질성을 어느 정도 확보해 보려는 의도에서였다.

이러한 기준에 따라 위에서 제시한 기간 중 도산한 전통적 기업과 전자상거래 기업의 수는 각각 22개와 19개였다. 같은 시기에 정상적으로 운영되던 동종업체 중에서 총자산 규모 및 종업원 수가 해당 도산 기업 수치의 80%~120% 이내인 31개의 전통적 기업과 27개의 전자상거래 기업을 비도산 기업으로 선정하였다. <표 1> 은 이들 자료의 기초통계량을 보여주고 있다.

<표 1> 표본 기업의 특성

단위: 억원

	일반기업			전자상거래기업		
	도산	비도산	합계	도산	비도산	합계
표본의 수	22	31	56	19	27	46
평균 자산 (표준편차)	183 (155)	192 (137)	188 (144)	147 (103)	163 (112)	156 (107)
평균 종업원 수 (표준편차)	173 (129)	168 (113)	170 (120)	123 (72)	132 (81)	128 (75)

### 4.2 연구변수의 선정

기업의 도산 예측에 관련된 문헌을 조사하여 56개의 재무비율을 선정하였다. 이 중 16개의 비율은 타 비율들의 부분적 또는 중복적 의미를 내재하고 있음으로 인해 제외시켰으며, 나머지 40개의 재무비율을 성장성, 수익성, 안정성, 활동성, 생산성 등 5개의 범주로 분류하였다. 연구의 활용된 연구 변수는 <표 2>와 같다.

<표 2> 연구변수의 구성 및 목표변수

구분	변수명	변수 설명	구분	변수명	변수 설명
성장성	B1	총자본증가율	안정성	D1	자기자본구성비율
	B2	유형고정자산증가율		D2	고정비율
	B3	유동자산증가율		D3	고정장기적합율
	B4	매출액증가율		D4	유동비율
	B5	순이익증가율		D5	당좌비율
수익성	C1	매출액총이익율		D6	부채비율
	C2	매출액영업이익율		D7	이자배율
	C3	매출액경상이익율		D8	현금흐름 대 부채비율
	C4	매출액순이익율		D9	현금흐름 대 총자본비율
	C5	총자본영업이익율	활동성	E1	총자본회전율
	C6	총자본경상이익율		E2	자기자본회전율
	C7	총자본순이익율		E3	재고자산회전율
	C8	자기자본경상이익율		E4	재고자산회전기간(일)
	C9	자기자본순이익율		E5	매출채권회전율
	C10	배당률		E6	매출채권회전기간(일)
	C11	배당성향		생산성	F1
	C12	주당순이익(천원)	F2		종업원1인당부가가치(천원)
	C13	주당영업현금(천원)	F3		종업원1인당경상이익(천원)
목표변수	'1' = 부도 '2' = 정상	F4	종업원1인당순이익(천원)		
		F5	자본집약도(천원)		
			F6		총자본투자효율
			F7		부가가치율

### 4.3 표본의 분할

표본의 양이 충분할 경우에는 일반적으로 표본을 모형추정용 (Training Data Set), 모형검증용 (Validation Data Set), 그리고 모형시험용 (Testing Data Set) 등으로 분류하여 모형을 구축, 검증, 시험한다. 그러나 본 연구에서는 채택된 표본의 양이 적기 때문에 잭나이프 (Jackknife) 방식을 사용하였는데 하나의 사례를 제외한 나머지 사례들로 모형을 추정하고, 추정된 모형을 토대로 앞서 제외된 사례를 예측하는 방식을 적용하였다. 그리고 같은 방식을 전체 사례의 수만큼 실행하여 모든 사례들이 모형추정시 한번씩 제외되어 예측되도록 하였으며, 개별 사례의 정확도를





## V. 연구결과의 의미

앞서 서술했듯이 의사결정나무 기법을 이용해 구축된 모형의 설명력은 다른 어떤 모형들보다 뛰어나다. 이 기법을 적용하여 기업의 도산 예측모형을 추정한 결과 전통적인 기업의 경우에는 ‘현금흐름 대 총자본비율’과 ‘부가가치율’이, 그리고 전자상거래 기업의 경우에는 ‘현금흐름 대 부채비율’, ‘총자본투자효율’, ‘유동비율’이 도산을 예측하는데 가장 높은 변별력을 지닌 변수로 지목되었다. 특히 두 모형에서 공히 지목한 ‘현금흐름’의 경우 Beaver (1996), Blum (1974a and 1974b), Deakin (1972 and 1977), Edminister (1972) 등도 자신들의 연구에서 기업 도산예측에 반드시 필요한 변수라 주장한 바 있다. 또한 Casey와 Bartczak (1984 and 1985)의 연구를 포함한 여러 자료에 따르면 ‘현금흐름’을 타 재무비율과 조합하여 사용할 경우 보다 정확한 기업 도산예측이 가능하다고 보고되고 있는데 이는 본 연구에서 제시하는 모형의 타당성을 뒷받침하고 있다.

기업의 성격에 따라 서로 다른 기업 도산예측 모형의 필요성을 강조하기 위하여 전통적인 기업과 전자상거래 기업을 위해 개발된 예측 모형을 전자상거래 기업의 자료에 적용하여 예측 결과를 비교하였다 <표 5>에서 보듯이 전통적인 기업 모형을 적용하는 경우 비도산 기업을 정분류하는 비율이 전자상거래 기업 모형에 비해 약 10% 정도 (90.7% → 80.6%) 떨어지는 반면, 도산 기업을 정분류하는 비율은 약 25% (70.4% → 45.2%)나 감소하고 있음을 알 수 있다. 이것은 전자상거래 기업의 표본 특성상 (19개의 도산기업과 27개의 비도산기업 사례로 구성) 임의로 도산기업을 예측 (Blind Guess) 하더라도 41.3% ( 19 / 46 ) 의 정확도를 기대할 수 있음으로 전자상거래 기업의 자료에 전통적 기업의 모형을 사용하는 것은 기업의 도산예측에 효과가 없음의 의미한다. 따라서 전자상거래 기업을 위한 또 다른 도산 예측모형의 필요성은 절실하다.

<표 5> 전자상거래기업에 적용한 모형별 예측력 비교

선택 모형	유형별 예측 정확도		전체 예측 정확도
	도산	비도산	
일반	45.2%	80.6%	66.0%
전자상거래	70.4%	90.7%	82.4%

이와 더불어 위에서 파악된 변수들이 국내 금융기관에서 기업의 신용평가를 위해 이용되는 변수들과 얼마나 일치하는가를 조사하였다. <표 6>은 은행 등의 금융기관에서 중소기업의 신용평가를 위해 일반적으로 고려하는 재무비율 항목들을 4가지 범주로 구분하여 분류하고 있는데 전통적 기업이나 전자상거래 기업의 모형에 포함되는 변수와 일치하는 항목이 하나도 존재하지 않는다는 것을 알 수 있다. 이같은 결과를 근거로 국내 금융기관들도 이제는 수년간의 영업활동을 통해 축적된 자료를 과학적으로 분석하여 의사결정에 활용해야 한다는 것을 지적하고자 한다.

&lt;표 6&gt; 중소기업 신용평가표 (재무상태 관련)

(출처: 월간 은행계)

구분	평가항목	배점	A	B	C	D	E	평점
재무상태	성장성	매출액증가율	3	3	2.4	1.8	1.2	0.6
	수익성	매출액영업이익율	5	5	4	3	2	1
		총자본경상이익율	5	5	4	3	2	1
	안정성	자기자본비율	5	5	4	3	2	1
		당좌비율	4	4	3.2	2.4	1.6	0.8
		부채비율	5	5	4	3	2	1
	활동성	총자본회전율	4	4	3.2	2.4	1.6	0.8
		영업자산회전율	4	4	3.2	2.4	1.6	0.8

## VI. 요약 및 결론

본 연구에서는 전통적인 기업의 도산 예측모형과 전자상거래 혹은 온라인 기업의 도산 예측모형을 개발하여, 기업도산 예측모형의 유동성을 살펴보고 모형의 변화에 영향을 미치는 다양한 요소들을 비교 분석함으로써 기업의 성격에 따른 새로운 모형의 필요성을 제시하였다. 의사결정나무 기법을 적용하여 기업의 도산 예측모형을 추정하고 분석한 결과 전통적인 기업의 경우에는 '현금흐름 대 총자본비율'과 '부가가치율'이, 그리고 전자상거래 기업의 경우에는 '현금흐름 대 부채비율', '총자본투자효율', '유동비율'이 도산을 예측하는데 가장 높은 변별력을 지닌 변수로 지목하였으며, 각 기업의 특성에 따른 새로운 모형의 필요성을 제시하였다.

본 연구에는 다음과 같은 연구의 한계성을 지니고 있는데 첫째, 연구결과를 일반화하기에는 자료의 양이 부족하다는 것이다. 둘째, 자료 수집의 어려움으로 인해 도산과 비도산 기업을 판별할 수 있는 다양한 변수들을 고려하지 못한 채 단지 재무비율만을 사용하였다. 보다 정확하고 포괄적인 모형을 위해서는 기업의 사업성, 거래신뢰도, 경영능력, 주식, 규모 등 다양한 변수들을 고려하여 모형의 변화를 관찰하는 것이 필요하다. 셋째, 비도산기업을 도산이라 잘못 예측하는 것보다는 도산할 기업을 비도산이라 예측하여 발생하는 비용이 더 클 수 있음에도 불구하고 오분류 비용을 동일한 것으로 가정한 채 모형을 추정함으로써 모형의 현실성이 간과되었다. 따라서 향후 정확한 오분류 비용을 반영하여 기존 모형을 재조정하는 작업이 요구된다. 마지막으로 전통적인 기업과 전자상거래 기업의 분류에 대한 한계이다. 앞에서도 언급하였지만 현재 100% 전통적 기업이거나 100% 전자상거래인 기업은 거의 없다고 할 수 있으며, 두 성격을 모두 가진 기업들이 대부분인 상황에서 기업의 분류는 큰 과제라 할 수 있다.

## 참고문헌

- 김영규와 감영규, 재무관리, 박영사, 2004.
- 이건창, “효과적인 의사결정을 위한 2 단계 하이브리드 인공지능경망 접근방법에 관한 연구, 경영정보학연구, 제5권, 제1호, 1995, pp. 36-51.
- 이건희, 재무관리, 학문사, 2002.
- 지청과 장하성, 재무관리, 법경사, 1996.
- Adam, Nabil, R., Dogramaci, Oktay, Gangopadhyay, Aryya, and Yesha, Yelena, *Electronic Commerce: Technical, Business, and Legal Issues*, Upper Saddle River: NJ, Prentice-Hall, 1999.
- Altman, E., "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy," *Journal of Finance*, Vol. 23, September 1968, pp. 589 - 609
- Altman, E., Haldeman, R., and Narayanan, P., "ZETA Analysis: A New Model to Identify Bankruptcy Risk of Corporations," *Journal of Banking and Finance*, Vol. 1, January 1977.
- Altman, E., *Corporate Financial Distress and Bankruptcy*, New York, NY: John Wiley and Sons, 1993.
- Altman, E., Macro, G., and Varetto, R., "Corporate Distress Diagnosis Comparisons Using Linear Discriminant Analysis and Neural Networks," *Journal of Banking and Finance*, Vol. 18, No. 3, 1994, pp. 505 - 529.
- Back, B., Laitinen, and Sere, K., "Neural Network and Genetic Algorithms for Bankruptcy Predictions," *Proceedings of the 3<sup>rd</sup> World Congress on Expert Systems*, 1996, pp. 123 - 130.
- Beaver, W., "Financial Ratios as Predictors of Failures. In Empirical Research in Accounting," selected studies supplement to the *Journal of Accounting Research*, Vol. 4, 1966, pp. 71 - 127.
- Berry, M., and Linoff, G., *Data Mining Techniques: For marketing, Sales, and Customer Support*, New York, NY: John Wiley and Sons, 1997.
- Billings, R., Milburin, T., and Schaalman, M., "A Model of Crisis Perception: A Theoretical and Empirical Analysis," *Administrative Science Quarterly*, Vol. 25, June 1980, pp. 300 - 316.
- Blum, M. P., "The Failing Company Doctrine," *Boston College Industrial and Commercial Review*, Vol. 16, 1974a.
- Blum, M. P., "Failing Company Discriminant Analysis," *Journal of Accounting Research*, Vol. 12, No. 1, 1974b.

- Brynjolfsson, Erik and Kahin, Brian (eds.), *Understanding the Digital Economy*, Boston, MA: MIT Press, 2000.
- Bryson, J., "A Perspective on Planning and Crises in the Public Sector," *Strategic Management Journal*, Vol 1, No. 2, April June 1981, pp. 181 - 196.
- Casey, C. and Bartczak, N., "Cash Flow, It's not the Bottom Line," *Harvard Business Review*, July-August 1984, pp. 60 - 66.
- Casey, C. and Bartczak, N., "Using Operating Cash Flow Data to Predict Financial Distress: Some Extensions," *Journal of Accounting Research*, Spring 1985, pp. 384 - 401.
- Chung, K. and Hong, K., "A Study on Bankruptcy Prediction Using Neural Networks," *Financial Management Study*, Vol. 12, No. 2, December 1995, pp. 1 - 23.
- Cisco Systems and the University of Texas at Austin, *Measuring the Internet Economy*, Report, January 2001.
- Deakin, E. B., "A Discriminant Analysis of Predictors of Business Failure," *Journal of Accounting Research*, Vol. 10, Spring 1972, pp. 167 - 179.
- Deakin, E. B., "Business Failure Prediction: An Empirical Analysis," In E. Altman and A. Sametz. (eds.) *Financial Crises: Institutions and Markets in a Fragile Environment*, New York, NY: John Wiley, 1977.
- Dutta, S. and Shekhar, S., "Bond Rating: A Non Conservative Application of Neural Networks," *IEEE Proceedings of the International Conference on Neural Networks*, 1988.
- Dutta, S., Shekhar, S., and Wong, W. Y., "Decision Support in Non Conservative Domains: Generalization with Neural Networks," *Decision Support Systems*, Vol. 11, No. 5, 1994, pp. 527 - 544.
- Edminster, R. O., "An Empirical Test of Financial Ratio Analysis for Small Business Failure Prediction," *Journal of Financial and Quantitative Analysis*, March 1972.
- Goldberg, D. E., *Genetic Algorithms in Search, Optimization and Machine Learning*, Reading, MA: Addison Wesley, 1989.
- Hunt, E., Martin, J., and Stone, P., *Experiments in Induction*, New York, NY: Academic Press, 1966.
- Kim, K., "Bankruptcy Sign and Prediction Based on Financial Ratios," *Korea Management Review*, Vol. 16, No. 2, February 1987, pp. 263 - 316.
- Klimasauskas, C., "Applying Neural Networks," In R. R. Trippi and E. Turban (eds.) *Neural Networks In Finance and Investing*, PROBUS, 1993, pp. 47 - 72.
- Lee, K., "Comparison of Accuracy on Corporate Bankruptcy: between MDA, Inductive Reasoning, and Neural Networks," *Korea Management Science*, 1993, pp. 123 - 134.

- Lee, K., "A Study on the Two Phased Hybrid Neural Network Approach to an Effective Decision Making," *The Journal of MIS Research*, Vol 5, No. 1, 1995, pp. 36 - 51.
- Lee, K., Kim, H., Kim, M., "An Inductive Learning Assisted Neural Network Approach to Bankruptcy Prediction: Comparison with MDA, Inductive Learning, and Neural Network Models," *Korea Management Review*, Vol. 24, 1995.
- Lee, K., Han, I., and Kwon, Y., "Hybrid Neural Network Models for Bankruptcy Predictions," *Decision Support Systems*, 1995.
- Lee, K. H., *Financial Management*. Seoul, Korea: Hakmoonsa, 1996.
- Michalewicz, Z., *Genetic Algorithms + Data Structures = Evolution Programs*, New York, NY: Springer Verlag, 1984.
- Norris, R. C., "Classifying High Performance Mutual Funds: A Comparison of C4.5, LDA & Logit," *Proceedings of the International Conference on INFORMS*, May 1996.
- Odom, M. D. and Sharda, R., "A Neural Network Model for Bankruptcy Prediction," *Proceedings of IEEE International Conference on Neural Networks*, 1990, pp. 151-173.
- OECD (Organization for Economic Co-operation and Development), *The Economics and Social Impact of Electronic Commerce*, 1999
- Ohlson, J., "Financial Ratios and the Probabilistic Prediction of Bankruptcy," *Journal of Accounting Research*, Vol. 18, No. 1, 1980, pp. 109 - 131.
- Platt, H. D. and Platt, M. B., "Development of Stable Predictive Variables: The Case of Bankruptcy Prediction," *Journal of Business Finance and Accounting*, Vol 17, No. 1, 1990, pp. 31 - 44.
- Platt, H. D. and Platt, M. B. A Note on the Use of Industry Relative Ratios in Bankruptcy Prediction. *Journal of Business Finance*, Vol. 1, No. 5:6, December 1991, pp. 1183-1194.
- Quinlan, J. R., "Induction of Decision Trees," *Machine Learning*, Vol. 1, 1986, pp. 81 - 106.
- Quinlan, J. R., *C4.5: Programs for Machine Learning*: Morgan Kaufmann, San Mateo, 1993.
- Quinlan, J. R., "Improved Use of Continuous Attributes in C4.5," *Journal of Artificial Intelligence Research*, Vol. 4, 1996a, pp. 77 - 90.
- Quinlan, J. R., "Bagging, Boosting, and C4.5," *Proceedings of 13<sup>th</sup> National Conference on Artificial Intelligence (AAAI 96)*, August 1996b.
- Raghupathi, W., Schkade, L. L., and Raju, B. S., "A Neural Network Approach to Bankruptcy Prediction," *Proceedings of the 24th HICSS*, 1991.
- Rahimain, E., Sing, S., Thammachote, T., and Virmani, R., "Bankruptcy Prediction by Neural Network," In R. R. Trippi and E. Turban (eds.) *Neural Networks In Finance and Investing*, PROBUS, 1993, pp. 159 - 176.

- Scott, J., "The Probability of Bankruptcy: A Comparison of Empirical Predictions and Theoretical Models," *Journal of Banking and Finance*, September 1981, pp. 317 - 344.
- Shin, K., Jo, H., and Han I., "Integration methodology of Multiple Techniques Using Genetic Algorithms: A Case of Corporate Prediction," *Proceedings of Korea MS/OR Conference*, 1996, pp. 199 - 202.
- Smart, C. and Vertinsky, I., "Designs for Crisis Decision Units," *Administrative Science Quarterly*, Vol. 22, December 1977, pp. 640 - 677.
- Smart, C. and Vertinsky, I., "Strategy and Environment: A Study of Corporate Responses to Crisis," *Strategic Management Journal*, Vol. 5, No. 3, July-September 1984, pp. 199 - 214.
- Song, I. M. and Yoon, S. S., *Intermediate Financial Accounting*, Seoul, Korea: Shinyoungsa, 1998.
- Stickney, C., *Financial Reporting and Statement Analysis: A Strategy Perspective*, New York, NY: The Dryden Press, 1996.
- Sung, S. and Han, J., "On Comparing Neural Networks and Discriminant Analysis: The Case of Business Failure Prediction," *Proceedings of KMIS Conference*, 1998, 12.1-12.12.
- Tam, K. and Chi, R., "Inducing Stock Screening Rules for Portfolio Construction," *Journal of Operations Research Society*, 1991.
- Tam, K. and Kiang, M., "Managerial Applications of Neural Networks: The Case of Bankruptcy Predictions," *Management Science*, Vol. 38, No. 7, July 1992, pp. 926 - 947.
- Weiss, L., "Bankruptcy Prediction: A Methodological and Empirical Update," *Working Paper*, Tulane, OK: Freeman School of Business, 1981.
- West, R. G., "A Factor Analytic Approach to Bank Condition," *Journal of Banking and Finance*, Vol. 9, No. 2, 1985, pp. 253 - 266.
- Westland, J. Christopher and Clark, Theodore H. K., *Global Electronic Commerce: Theory and Case Studies*, Cambridge, MA, The MIT Press, 1999.
- White, G., Sondhi, A., and Fried, D. *The Analysis and Use of Financial Statements*. New York, NY: John Wiley and Sons, 1994.
- Wilcox, J. A., "Simple Theory of Financial Ratios as Predictors of Failures," *Journal of Accounting Research*, Autumn 1991.
- Wilson, R. L. and Sharda, R., "Bankruptcy Prediction Using Neural Networks," *Decision Support Systems*, Vol. 11, No. 5, 1994, pp. 545 - 557.
- Zavgren, C. V., Dugan, M. T., and Reeve, J. M., "The Association between Probabilities of Bankruptcy and Market Responses: A Test of Market Anticipation," *Journal of Business Finance and Accounting*, Spring 1988, pp. 19 - 45.
- Zmijewski, M. E., "Methodological Issues Related to the Estimation of Financial Distress Prediction Model," *Journal of Accounting Research*, Vol. 22, 1984, pp. 59 - 82.

<Abstract>

## An Application of Data Mining Techniques in Electronic Commerce

Tae-Kyung Sung · Seok-Chin Chu · Joong-Han Kim · Jun-Seok Hong

This paper uses a data mining approach to develop bankruptcy prediction models suitable for traditional (off-line) companies and electronic (on-line) companies. It observes the differences in the composition prediction models between these two types of companies and provides interpretation of bankruptcy classifications. The bankruptcy prediction models revealed the major variables in predicting bankruptcy to be 'cash flow to total assets' and 'gross value-added to net sales' for traditional off-line companies while 'cash flow to liabilities', 'gross value-added to net sales', and 'current ratio' for electronic companies. The accuracy rates of final prediction models for traditional off-line and electronic companies were found to be 84.7%, and 82.4%, respectively. When the model for traditional off-line companies was applied for electronic companies, prediction accuracy dropped significantly in the case of bankruptcy classification (from 70.4% to 45.2%) at the level of a blind guess (41.30%). Therefore, the need for different models for traditional off-line and electronic companies is justified.

**Keywords** : Data Mining, Bankruptcy Prediction Model, On-line Company

\* 이 논문은 2005년 10월 4일 접수하여 1차 수정을 거쳐 2005년 11월 3일 게재 확정되었습니다.