

데이터 마이닝을 이용한 서울시 오피스빌딩 투자특성 예측에 관한 연구*

홍아름** · 고재풍*** · 유선중****

A Study on the Forecasting Model of the Investment Characteristics of Seoul Office Buildings based on Data Mining*

Areum Hong** · Jaipoong Ko*** · Seonjong Yoo****

요약 : 2001년 부동산 투자회사법의 제정으로 다양한 형태의 간접투자가 가능하게 되었다. 특히 오피스 빌딩이 안정적이고 높은 수익을 기대할 수 있는 투자처라는 인식으로 오피스빌딩에 대한 간접투자가 지속적으로 늘어나고 있다. 그러나 간접투자에 관한 연구는 간접투자된 오피스빌딩의 유형에 관한 데이터의 양이 많지 않고, 그 데이터의 수집마저 어렵기 때문에 기존 연구건수는 많지 않다. 하지만 본 연구는 양질의 데이터로 투자방식에 관한 실증분석을 하였다는 점에서 의의가 있다. 본 연구는 1998년부터 2009년 1/4분기까지 16,500㎡ 이상의 서울시 오피스빌딩을 대상으로 부동산 간접 투자제도가 도입된 이후 서울시 오피스빌딩 시장에서의 직접투자와 간접투자를 결정하는 요인을 데이터 마이닝을 이용하여 분석하였다. 오피스빌딩을 매입 방식별로 나누어 투자자가 선호하는 요인을 분석하였고, 그 요인을 분석하기 위해 로지스틱 회귀분석, 의사결정나무 분석, 신경망 분석을 사용하였고, 이 세 가지 분석방법의 예측력을 비교, 평가하였다.

주제어 : 오피스빌딩, 간접투자, 예측력, 데이터 마이닝

ABSTRACT : Enactment of 'Korean REITs Law and Asset Management Act' in 2001 has triggered numerous types of investments in securitized real estate (indirect real estate investments). In particular, the securities with office properties as underlying assets have grown substantially since office properties are expected to yield stabilized high returns. However, the restrictive number of securitization of office properties and the difficulty of data collection have posed challenges to studying the securitization of real estate in Korea. Nonetheless, to investigate what variables affect the investment decision of securitized real estate, this study analyzes the reliable date of office properties, larger than 16,500㎡, in Seoul for the period between 1998 Q1 to 2009 Q1. Logistic regression, decision tree, and neural network are employed and compared for their respective predictive capabilities.

Key Words : office building, indirect investment, predictability, data mining

* 이 논문은 2010학년도 건국대학교의 지원에 의하여 작성되었습니다.

** 건국대학교 대학원 부동산학과 석사과정 수료(Master, Dept. of Real Estate, Konkuk Univ.)

*** 건국대학교 부동산대학원 겸임교수(Lecturer, Dept. of Real Estate, Konkuk Univ.)

**** 건국대학교 부동산학과 부교수(Associate Professor, Dept. of Real Estate, Konkuk Univ.),

교신전자(E-mail: yoojs@konkuk.ac.kr, Tel: 02-2049-6063)

I. 서론

외환위기로 기업 및 금융기관들의 자금유동성이 악화되었다. 일시적으로 환금성이 현저하게 낮아진 부동산을 신속하게 처리하기 위해 외국인 투자자들에게 저평가된 오피스빌딩을 매각할 수밖에 없었다. 외국인 투자가 허용되면서 부동산에 대한 간접투자의 길이 열렸고 국내외 자본에 의한 부동산 간접투자가 활발히 이루어지고 있다. R2Korea(2009)의 발표에 따르면 리츠의 경우 2009년 3분기 말, 전분기 대비 1조 6,707억 원 증가하였으며, 부동산 관련 펀드는 2009년 10월 말, 6월 대비 1조 2,298억 원이 증가하였다.

간접 투자시장은 꾸준히 증가하고 있지만 간접 투자에 관한 연구는 외국의 간접 투자제도의 비교 분석이나 우리나라 간접 투자제도의 문제점 및 활성화 방안 등을 연구하는 데 그치고 있다. 그것은 간접투자에 관한 연구는 간접투자된 국내외 자본이 선호하는 오피스빌딩의 유형에 관한 데이터의 양이 많지 않고, 그 많지 않은 데이터의 수집마저 어렵기 때문이다.

본 연구에서는 간접 투자제도가 도입된 이후 서울시 오피스빌딩 시장에서 간접투자자와 직접투자¹⁾를 결정하는 요인을 분석하고, 그 요인분석을 위해 로지스틱 회귀분석, 의사결정나무 분석, 신경망 분석을 하였으며, 이 세 분석모형에 대한 예측력을 비교, 평가하는 데 그 목적이 있다.

본 연구는 자료구득이 가능한 1998년 이후부터 2009년 1분기까지 16,500㎡ 이상인 서울시 오피스빌딩 총 147건을 대상으로 하였다. 분석자료에 사용

된 오피스빌딩의 자료는 건축물대장과 등기부등본을 발급받아 소유법인과 기타 정보들의 확인을 거쳐 조사된 (주)신영에셋의 거래사례 자료이다. ABS 자산의 경우 금융감독원의 전자 공시시스템을 사전에 확인하였고, 리츠와 부동산펀드 자산의 경우 공시내역을 확인하였다. 그리고 거시적 특성변수들의 경우 한국은행의 분기별 통계 자료를 조사하였다.

간접투자 관련 선행연구를 보면 김선주·이춘섭(2005)은 주식, 국채, 오피스빌딩의 포트폴리오 분석을 통해 부동산 펀드의 자산운영에 관한 분석을 하였고, 강혜진·오동훈(2009)은 간접투자된 오피스빌딩의 실체를 분석하고 입지성, 기능성, 시장성 등 세부특징을 조사하여 수익률과의 상관관계를 분석하였다. 이처럼 간접투자된 오피스빌딩에 대한 실증분석 연구는 많지 않다.

본 연구에서는 오피스빌딩을 매입 방식별로 나누어 투자자가 선호하는 요인을 분석하였고, 그 요인을 분석하기 위해 로지스틱 회귀분석, 의사결정나무 분석, 신경망 분석을 사용하였고, 이 세 가지 분석방법의 예측력을 비교 평가하였다. 로지스틱 회귀분석을 통해 간접 투자방식을 선택하기 위해 어떠한 정보를 수집하고 관리하는 것이 유용한지를 예측하였으며, 의사결정나무분석을 통해서는 로지스틱회귀분석으로는 한계가 있는 각 입력변수의 영향이 다른 입력변수에 어떠한 영향을 미치는지에 대해 알고자 하였다.

신경망 분석은 매우 유연하기는 하지만 결과를 해석하는 것이 매우 어렵다. 즉, 로지스틱 회귀분석이 제공하는 것과 같은 계수들에 대한 간편한 해석이 불가능하여 어떤 입력변수가 중요한지 또

1) 외국인 투자란 외국인(외국 국적을 가진 자연인 또는 외국의 법률에 의거하여 설립된 법인)이 자국 이외의 나라에 투자하는 일을 말한다. 외국인 투자는 증권투자의 형태를 많이 띠는데, 시장을 경유하지 않고 자본을 투자하는 직접투자와 시장을 경유하여 주식·채권을 취득하는 간접투자자로 분류된다. 본 논문에서는 외국자본이 오피스빌딩을 취득 또는 매각에 있어서 대주주의 역할을 하거나, 취득 또는 매각과 관련된 의사결정을 하는 데 중요한 역할을 하는 경우는 직접투자자로 분류하였고, 그렇지 않은 경우는 간접투자자로 구분하였다.

는 그것들이 어떻게 상호작용하는지를 파악하기 어렵다. 신경망 분석의 경우 비선형성을 갖는 변수에 유용하고 예측의 정확성은 높지만 분류나 예측결과만 제공할 뿐 과정에 대한 설명이 어렵다. 이러한 단점을 로지스틱 회귀분석과 의사결정나무 분석을 통해 보완하였다. 로지스틱 회귀분석은 간접투자자 및 직접투자자를 이항 종속변수로 로지스틱 회귀분석을 하였다.

간접투자자와 직접투자자를 종속변수로 한 로지스틱 회귀분석을 통해서 얻는 독립변수의 계수로 해석측면의 이해와 결론을 도출하였으며, 독립변수들 간의 영향력을 파악하기 위해 의사결정나무 분석을 통해 분석하였다.

관련된 선행연구를 살펴보면 홍종선·박용석(2005)은 로지스틱 회귀분석과 의사결정나무 분석을 통하여 기업의 부실예측 모형을 구축하여 수익성, 활동성, 성장률, 유동성, 안정성을 나타내는 변수로 금융기관의 통계적 신용에 대해 평가하여 그 효율성을 분석하였다. 두 분석의 결과 의사결정나무 분석이 정확도, 예측력, 안정성에서 우수

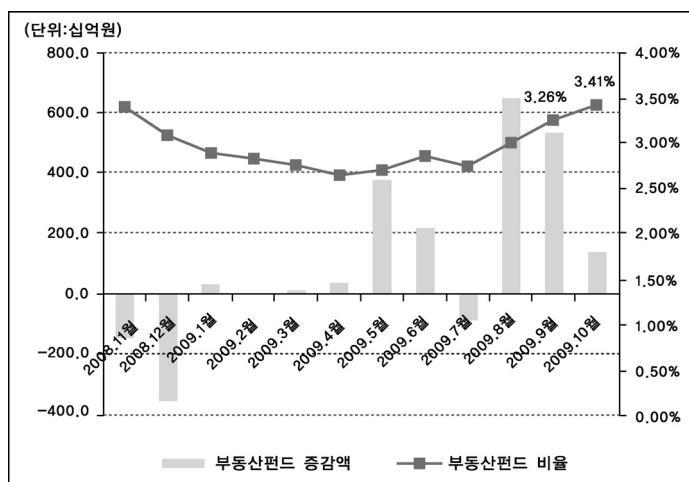
하다는 결론을 내렸다. 또한 이준용 외(2008)는 다중 회귀분석과 의사결정나무 분석, 신경망 분석을 통하여 아파트 가격을 예측하였다. 그 결과 강남권 아파트 가격 결정요인과 그 가격을 결정하는 요인 간의 상호관계를 파악하였으며, 세 모형을 통해 아파트 가격 모형 구축에 대한 가능성을 제시하였다.

II. 간접투자시장 현황 및 자료 분석

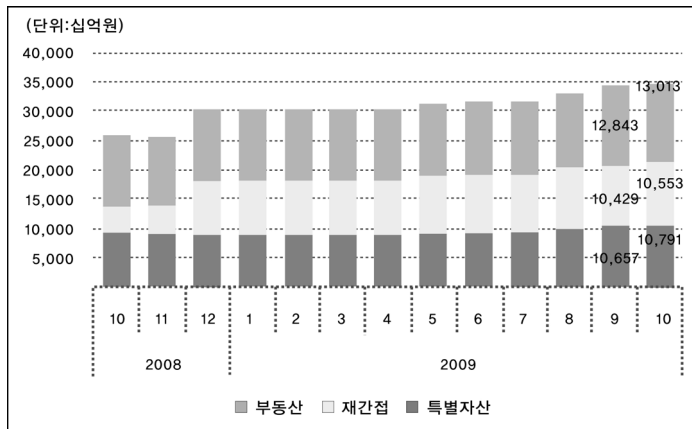
1. 간접투자시장 현황

1) 부동산 관련 펀드

R2Korea(2009)에 따르면, 2009년 1분기, 전체 펀드 설정액은 37조 7,839억 원이며, 부동산 관련 펀드는 총 3조 3,626억 원으로 전체 펀드 설정액의 8.9% 수준이며, 전분기 대비 1,096억 원이 증가한 것이다. 상품별로는 부동산 9,013억 원, 재간접 투자자산 1조 2,152억 원, 특별자산 1조 2,461억 원을 차지한다.

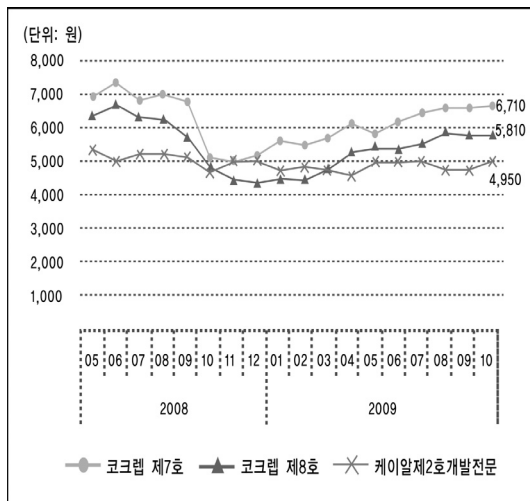


〈그림 1〉 부동산 자산 총액 증감 및 비율



〈그림 2〉 월별 부동산펀드 설정액

2) REITs 현황



〈그림 3〉 리츠 주가동향

2009년 3분기 말, 리츠의 총자산은 전분기 대비 1조 6,707억 원 증가한 6조 1,403억 원 규모며, 코

람코자산신탁의 자산규모는 2조 5,894억 원 규모로 전체의 37%를 차지한다. 리츠의 주가동향은 큰 폭의 변동 없이 2009년 2분기와 1분기가 유사한 수준이다.

2. 자료 분석

분석자료²⁾를 통해 서울시에 소재한 16,500㎡ 이상 오피스빌딩 총 147건을 대상으로 보면, 먼저 권역의 경우 도심권역의 거래사례가 39.5%로 가장 높게 조사되었으며, 직접투자된 경우 강남권역의 오피스빌딩 거래사례가 38.4%로 가장 높았으며, 간접투자자의 경우 도심권역이 48.6%로 가장 높았다. 이로부터 투자방식에 따라 권역별 선호도가 다른 것으로 조사되었다. 연면적은 직접투자된 경우 평균 43,777㎡, 간접투자된 경우 평균 51,096㎡

2) 데이터 마이닝은 자료(data)에 저장되어 있는 방대한 양의 자료로부터 의사결정에 도움이 되는 유용한 정보를 발견하는 일련의 작업들의 집합을 의미한다. 즉, 데이터 마이닝은 “데이터에 함축되어 있으나 이전에는 발견되지 않은 잠재적으로 유용한 정보를 새롭게 추출”함으로써 과거행위의 분석을 기초로 미래 행위를 예측하는 의사결정지원을 위한 모형을 만들어 내는 것이라고 할 수 있다. 오피스빌딩의 경우 거래가격이나 임대료, 영업경비와 같은 자료는 소유자의 영업비밀로 취급되어 공개를 꺼리는 시장분위기와 간접투자된 국내외 자본이 선호하는 오피스빌딩의 유형에 관한 데이터 수집이 어렵다. 그렇지만, 비록 관측치가 적어 의미 있는 데이터 마이닝을 도출하기 어렵다 해도 16,500㎡ 이상의 서울시 오피스 매매자료에 대한 10년간의 데이터를 구축하면서 데이터 마이닝을 시도하게 되었다. 이러한 시도는 김상용(2009), 이준용(2008) 등의 선행연구에서도 볼 수 있다.

〈표 1〉 기술통계량

(단위: 건수, 분, m², %)

구분		직접투자		간접투자		Total	
		n	%, mean	n	%, mean	n	%
권역	CBD	22	30.1%	36	48.6%	58	39.5%
	KBD	28	38.4%	25	33.8%	53	36.1%
	YBD	19	26.0%	10	13.5%	29	19.7%
	ETC	4	5.5%	3	4.1%	7	4.8%
지하철 접근성		73	5.9	74	5.1	147	100%
연면적		73	43,777	74	51,096	147	100%
거래면적		73	40,817	74	41,008	147	100%
노후연도		73	11.8	74	10.7	147	100%
외국자본		9	12.3%	48	64.9%	57	38.8%
국내자본		64	87.7%	26	35.1%	90	61.2%
Total		73		74		147	100%

로 간접투자된 경우가 평균 7,319m² 넓은 것으로 나타났다. 또한 평균 거래면적은 간접투자된 경우 41,008m²로 직접투자된 경우가 40,817m²보다 약간 넓었다. 국내자본과 외국자본을 보면 국내자본이 61.2%로 대부분을 차지하고 있다. 특히 87.7%의 국내자본은 직접투자되고, 64.9%의 외국자본이 간접투자되고 있는 것으로 분석되었다.

III. 실증분석 및 결과

1. 검증모형

데이터 마이닝의 경우 모형의 타당성 평가를 위해 데이터를 분석용 데이터, 검증용 데이터, 테스트 데이터로 분할한다. 분석용 데이터는 모형을 구축하는 데 이용하고, 검증용 또는 테스트 데이터는 최종적인 평가를 수행하는 데 이용한다.

- ① 분석용 데이터(Training Data): 데이터를 분석하여 모형을 만드는 데 직접적으로 사

용되는 데이터이다.

- ② 검증용 데이터(Validation Data): 모형의 성능을 감독하고 개선하기 위해 간접적으로 사용되는 Tree노드의 가지치기, Regression노드의 변수선택, Neural Network노드의 수렴값 결정 등에 사용된다.
- ③ 테스트 데이터(Test Data): 모형의 생성에 전혀 사용되지 않으며, 일반화의 검토를 위해 남겨두는 데이터이다.

본 연구에서는 연습용 데이터 80%, 검증용 데이터 10%, 테스트 데이터 10%로 partition을 나눴는데 이는 데이터의 양이 많지 않아 연습용 데이터로 분석한 결과의 유의성이 떨어지기 때문이다.

1) 로지스틱 회귀분석

본 연구에는 이항 로지스틱 회귀분석을 사용하였다. 이는 사건이 일어날 경우의 수가 두 가지인 모형을 말한다.

로지스틱 회귀분석은 확률에 대한 로짓변환을 고려하며 식 (1)과 같다.

$$\ln\left(\frac{P_y}{1-P_y}\right) = \alpha + B_1X_1 + B_2X_2 + \dots + B_pX_p \quad (1)$$

설명변수가 분류 결정에 미치는 영향의 정도는 *Odds Ratio*로 계량화할 수 있으며 식 (2)와 같다.

$$\begin{aligned} Odds(y=1) &= \frac{\hat{P}(y=1|X_1, X_2, \dots, X_p)}{1 - \hat{P}(y=1|X_1, X_2, \dots, X_p)} \\ &= \text{EXP}(\alpha + B_1X_1 + \dots + B_pX_p) \end{aligned} \quad (2)$$

여기서 *Odds Ratio*는 어떤 사건이 발생되지 않을 확률과 비교해서 발생될 확률의 비율을 의미한다. 정광모 외(2003)는 *Odds Ratio*의 변화는 $\text{EXP}(B)$ 배가 되는데 1보다 크면 *Odds Ratio*가 $\text{EXP}(B)$ 배로 커지고 1보다 작으면 *Odds Ratio*가 $\text{EXP}(B)$ 배만큼 작아진다고 해석한다.

2) 의사결정나무 분석

의사결정나무(Decision Tree, Answer Tree) 분석은 의사결정 규칙을 나무구조로 도표화하여 분류와 예측을 수행하는 방법으로 모형을 사용하는 자가 쉽게 이해할 수 있고, 두 개 이상의 변수가 결합한 상호작용의 효과가 반응변수에 어떤 영향을 주는지 쉽게 알 수 있다.

의사결정나무는 분리기준(Split Criterion), 정지규칙(Stopping Rule) 등을 분석의 목적과 자료 구조에 따라 적절히 지정하여 얻어진다.

분리기준은 하나의 부모마디로부터 자식마디들이 형성될 때 입력변수(Input Variable)의 선택과 범주(Category)의 병합이 이루어질 기준을 말하

는데 본 연구에서는 분리기준으로 카이스퀘어 검정을 사용하였다.

또한 정지규칙은 더 이상 분리가 일어나지 않고 현재의 마디가 끝마디가 되도록 하는 여러 가지 규칙을 의미한다. 예측 오차값을 줄여나가면서 가지치기를 하며, 오차값이 변화가 없을 때 가지치기를 중지하는 것이다.

본 연구에서는 다음과 같이 정지규칙을 설정하였다.

- ① 분리된 마디는 최소한 2개 이상의 관측치를 가져야 한다.
- ② 마디의 관측치가 5개 이하일 경우 더 이상 분리하지 않는다.
- ③ 나무 깊이의 최댓값은 6이다.
- ④ 검정 시 p값이 0.2 이상일 경우에는 더 이상 분리하지 않는다.

3) 신경망 분석

신경망(Artificial Neural Networks)에 관한 연구는 뇌신경생리학으로부터 영감을 얻어 복잡한 구조를 가진 데이터에서의 예측문제를 해결하는데 유용한 분석모형이다. 신경망은 매우 유연하기는 하지만 로지스틱 회귀분석처럼 계수들을 통한 해석이나 의사결정나무 분석처럼 어떤 입력변수가 어떤 입력변수에 상호영향을 미치는지를 해석하기는 매우 어렵다.

본 연구에서는 데이터 분석을 위해 가장 널리 사용되는 MLP(Multilayer Perceptron, 다층인식자) 신경망을 사용하였다.

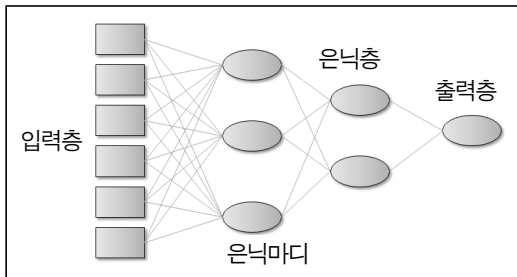
MLP는 입력층, 인간의 신경세포를 모형화한 은닉마디, 은닉마디로 구성된 은닉층 그리고 출력층으로 이루어졌다.

- ① 입력층(Input Layer): 각 입력변수에 대응하는 마디들로 구성되어 있다. 명목형 변수에 대해서는 각 수준에 대응하는 입력마디를 가지

게 되고 이는 더미변수로 사용하는 것과 같다.

② 은닉층(Hidden Layer): 여러 개의 은닉마디로 구성되어 있다. 각 은닉마디는 입력층으로부터 전달되는 변수값들의 선형결합을 비선형함수로 처리하여 출력층이나 다른 은닉층에 전달한다.

③ 출력층(Output Layer): 목표변수에 대응하는 마디들을 갖는다.



〈그림 4〉 MLP(Multilayer Perceptron)의 구조

MLP의 구조를 도식화하면 식 (3)과 같다.

$$\begin{aligned} H_1 &= f_1(b_1 + w_{11}X_1 + w_{21}X_2 + \dots + w_{p1}X_p) \\ H_2 &= f_2(b_2 + w_{12}X_1 + w_{22}X_2 + \dots + w_{p2}X_p) \\ Y &= g(b_0 + w_{10}H_1 + w_{20}H_2) \end{aligned} \quad (3)$$

(f : 쌍곡탄젠트, g : 로지스틱)

4) 모형의 평가

모형평가는 관심 성과치의 예측을 위해 만든 모형이 임의의 모형보다 우수한지, 고려된 서로 다른 모형들 중 어느 것이 우수한 예측력을 보유하고 있는지를 비교분석하는 과정이다.

Lift 그래프에서는 누적%Response값을 보게 되는데, 이는 다음 식 (4)와 같다.

$$\text{누적\%Response} = \frac{\text{목표변수의 특정범주 빈도}}{\text{목표변수의 전체빈도}} \quad (4)$$

따라서 상위집단에서 누적%Response값이 높으면 상위집단에서 특정범주의 확률이 높은 것을 보여주며 이는 곧 간접투자하는 경우와 직접투자하는 경우와의 변별력을 높여준다는 것을 의미한다.

Response Threshold 그래프에서는 정확도를 보게 되며 이는 식 (5)와 같다.

$$\text{정확도(accuracy)} = \frac{\text{실제로 특정범주에 포함되는 빈도}}{\text{특정범주에서 분류기준값 넘는 빈도}} \quad (5)$$

이 도표를 작성하는 것은 분류기준값이 달라짐에 따라 모형의 분류결과가 얼마나 변동이 없는가를 알기 위해서이며, 이때 분류기준값의 변화에 따른 모형 분류결과와 변동여부를 모형의 일치성이라고 한다. 분류기준값이 분류결과에 영향을 주지 않는 것이 좋은 모형이므로 정확도가 극단으로 존재하는 것이 조금씩 변화하는 것보다 좋은 모형이라고 할 수 있다. 그러나 사후확률이 양극단으로 존재한다고 해도 실제 반응변수의 범주값과 일치하는 결과를 가져오지 않는다면 문제가 발생하기 때문에 높은 사후확률에 대해서 실제 반응변수의 특정 범주값의 빈도가 높은 것을 좋은 모형으로 판단한다.

또한 ROC 그래프를 통해 민감도와 특이도를 계산할 수 있다. 민감도란 연구자가 관심을 가진 데이터를 가장 잘 분류하는 능력을 나타내는 척도를 말하며, 특이도란 연구자의 관심 밖의 데이터를 잘 분류하는 능력을 나타내는 척도이다. 이는 식 (6)과 식 (7)을 통해 알 수 있다.

민감도(*sensitivity*)

$$\begin{aligned} & \text{(실제 : 간접투자, 예측 : 간접투자)인} \\ & = \frac{\text{관찰치의 빈도}}{\text{실제 : 간접투자인 관찰치의 빈도}} \end{aligned} \quad (6)$$

특이도(*specificity*)

$$\begin{aligned} & \text{(실제 : 직접투자, 예측 : 직접투자)인} \\ & = \frac{\text{관찰치의 빈도}}{\text{실제 : 직접투자인 관찰치의 빈도}} \end{aligned} \quad (7)$$

ROC 그래프를 보면 민감도와 특이도의 상관관계를 알 수 있게 된다. 이 그래프에서 대각선은 우연에 의한 ROC 곡선을 의미한다. 즉, 예측이 전혀 없는 모형에 대한 ROC 곡선을 의미한다. 강현철 외(2009)에 따르면, 좋은 성능을 가진 모형일수록 ROC 곡선은 대각선 위쪽에 위치하게 된다.

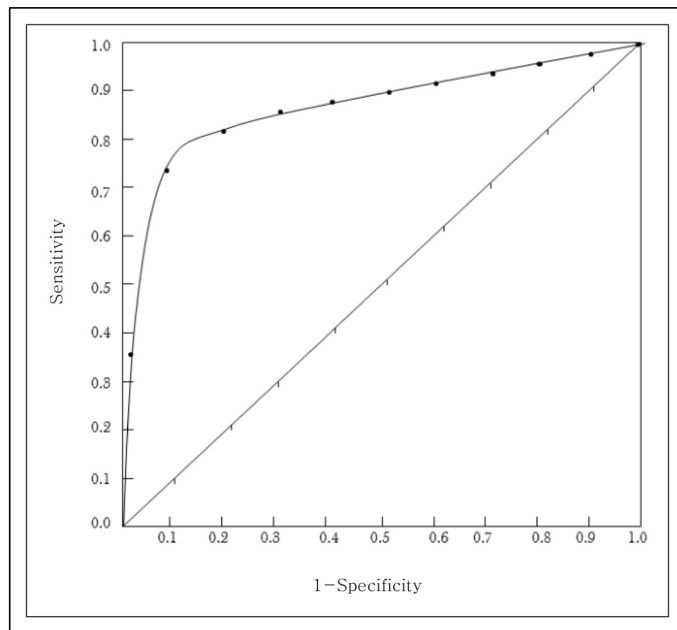
2. 변수 설정

1) 종속변수

본 연구는 간접투자자와 직접투자자로 구분하여 각 투자구조로 매입된 서울시 오피스빌딩의 특성의 차이를 분석하였다. 이때 ABS, REITs, REF를 '간접투자'로 더미변수로 만들어 그 값을 1로 선택하고, '직접투자'는 0으로 선택하였다.

2) 독립변수

매입방식에 영향을 미치는 독립변수로 '권역'은 권역별로 임대료, 관리비, 공실률 등의 차이가 있어 지리적 차이가 오피스빌딩의 매입방식에 영향을 줄 수 있을 것으로 예상하였다. 따라서 서울시의 '3대 권역(CBD, KBD, YBD)'과 '서울 기타 권역(ETC)'의 4개 지역을 더미변수로 만들었다. 권역은 이상경·이인철(2005), 이준용(2008)의 연구에서 외국자본은 도심권역, 여의도 권역에서 간접



〈그림 5〉 ROC 그래프

투자하는 것을 선호한다고 하였기에 변수로 선택하였다.

‘지하철역 접근성’은 일반적으로 임차인이 사무실을 사용하는 데 있어 교통적 요건이 우수한 오피스빌딩을 선호하고, 간접투자자와 직접투자자와의 차이가 있는지 분석하기 위해 도보로 분 단위로 측정하여 변수로 선택하였다.

본 연구에서는 ‘연면적’과 ‘거래면적’을 분리하여 분석하였다. 투자자 특성에 따라 전체 매입과 부분 매입에 대한 선호도가 다를 것이라고 예상되었기 때문이다. 또한 김상용(2009)의 선행연구를 따라 ‘거래면적’은 최근에 층별 매각이나 상가를 제외한 오피스 일부 매각 사례가 증가하는 추세이기 때문에 다수의 투자자가 투자하는 간접투자자와 직접투자자와의 차이가 있는지를 분석하기 위해 변수로 선택하였다.

오피스빌딩의 ‘노후년도’는 오피스빌딩을 준공한 시점의 준공연도를 거래시점에서 차감한 값이다. 간접투자 주체의 경우 자산운용을 통해 발생

하는 현금흐름에 의존하기 때문에 감가상각이 높은 노후화된 오피스빌딩의 매입방식에 영향을 줄 수 있을 것으로 예상하고 변수로 선택하였다.

외국자본은 각 매입방식을 통한 투자경험과 노하우가 축적되어 있기 때문에 관련 법에 맞추어 매입방식을 변화시켜 조세부담을 최소화하고 투자수익을 극대화했다. 외국자본과 달리 국내자본의 경우 국내시장에서 각 방식을 도입하기에는 각 매입방식에 대한 경험이 부족하고 투자자 확보를 위한 시장이 활성화되지 않았기 때문에 관련 법 제정에 따라 바로 매입방식의 변화가 이루어지지 않았다. 그러므로 이준용(2008)과 같이 ‘국내자본’이나 ‘외국자본’이 오피스빌딩의 매입방식에 차이를 주었을 것으로 예상하고 변수로 선택하였다.

오피스빌딩을 투자목적으로 매입하거나 재매각하기 위해서는 무엇보다도 매입금액이 중요하기 때문에 ‘㎡당 금액’을 변수로 선택하였고, ‘월 임대료’와 ‘월 관리비’는 오피스빌딩의 운영수익을 결정하는 중요한 요소이기 때문에 변수로 선택하여

〈표 2〉 모형에 사용된 변수

구분	변수	내용	단위
종속변수	투자구조 (base = 직접투자)	간접투자(ABS, REITs, REF)	1
		직접투자	0
독립변수	권역별 (base = ETC)	CBD	1
		KBD	1
		YBD	1
		ETC	0
	지하철 접근성	지하철역에서부터의 도보 거리	분
	연면적	건물 각 층의 바닥면적 합계	㎡
	거래면적	실제 매매 당시의 거래면적	㎡
	노후년도	거래연도-준공연도	년
	매입자본 (base = 국내자본)	외국자본	1
		국내자본	0
	월 임대료	매입시점의 ㎡당 월 임대료	천원
	월 관리비	매입시점의 ㎡당 월 관리비	천원
	㎡당 금액	단위면적(㎡)당 가격	천원
	환율	원/미국달러(기준환율)	원
	회사채	장외 3년, AA-등급	%

분석하였다.

R2Korea(2009)의 시장동향보고서를 보면 주거지수나 시장금리, 주요 경제지표 등을 고려하여 오피스빌딩의 시장동향을 살피고 있다. 특히 〈표 1〉 기술통계량에서 보이는 바와 같이 간접투자의 경우 외국자본으로의 투자가 많기 때문에 거시적 특성들이 영향을 줄 것으로 예상하고 ‘회사채’와 ‘환율’을 선택하여 분석하였다.³⁾

3. 분석결과

서울시 오피스빌딩에 투자할 때 간접투자에 영향을 미치는 요인들을 분석하기 위해 로지스틱 회귀분석을 사용하였다. 로지스틱 회귀분석을 통해 얻

어진 각 변수들의 계수의 영향력 해석을 위해 의사결정나무 분석을 사용하였다. 또한 로지스틱 회귀분석, 의사결정나무 분석과 신경망 분석의 세 분석들의 장점과 단점을 상호 보완하는 모형분석을 하였으며, 세 분석의 정확도, 민감도, 특이도를 비교분석하여 가장 예측력이 높은 모형을 선별하였다.

본 연구에서는 모형의 타당성 평가를 위해 Training Data 80%, Validation Data 10%, Test Data 10%로 설정하여 분석하였다.

1) 로지스틱 회귀분석 결과

간접투자에 영향을 미치는 다양한 독립변수들을 대상으로 로지스틱 회귀분석을 하여 다음과 같은 결과가 도출되었다.

〈표 3〉 로지스틱 회귀분석 결과

Parameter		B	S. E.	Wald	Exp(B)
권역	CBD	0.40	0.68	0.34	1.49
	KBD	0.28	0.61	0.20	1.32
	YBD	0.44	0.65	0.45	1.55
지하철 접근성		0.01	0.07	0.04**	1.02
연면적		0.00	0.00	3.32*	1.00
거래면적		-0.00	0.00	5.06**	1.00
노후년도		-0.02	0.04	0.29	0.98
외국자본		1.71	0.36	22.35***	5.52
월 임대료		0.27	0.15	3.30*	1.32
월 관리비		-0.43	0.48	0.80	0.65
㎡당 금액		-0.00	0.00	0.09	1.00
환율		-0.00	0.00	4.26**	1.00
회사채		-0.42	0.21	3.91**	0.66
상수		8.82	3.79	5.41**	999.00

N: 118

종속변수: 직접투자(0), 간접투자(1)

Likelihood Ratio 67.08***

주: *, **, ***는 각각 10%, 5%, 1% 유의수준에서 유의

3) 독립변수간의 공선성에 대한 진단은 변수선택 단계에서부터 분석과정별로 VIF, 공차한계, 상태지수 등을 검토하였다.

전체 147건 분석 결과 유의성 검정(Likelihood Ratio Test)에 대해 유의수준 1% 하에서 유의한 것으로 나타났다.

독립변수의 상대적인 영향력은 *Odds Ratio*로 표현할 수 있다. 이를 적용하여 분석해 보면 먼저, 도보로의 '지하철 접근성'이 1분 증가할 때 직접투자 주체가 매입할 *Odds Ratio*보다 간접투자자로 매입할 *Odds Ratio*가 1.02배 증가하였는데 이는 직접투자자와 거의 차이가 없는 것으로 판단된다.

'거래면적'의 경우 1㎡ 증가할 때 직접투자 주체가 매입할 *Odds Ratio*보다 간접투자 주체가 매입할 *Odds Ratio*가 1배로 직접투자자와 차이가 없었다.

'외국자본'을 보면, 국내자본으로 간접투자할 경우보다 '국내자본'으로 간접투자 주체가 매입할 *Odds Ratio*가 5.52배 증가하고 있는데 이는 예상했던 바와 같이 국내자본의 경우 국내시장에서 각 매입방식을 도입하기에는 그에 대한 경험의 부족하고, 투자자 확보를 위한 시장이 활성화되지 않았기 때문인 것으로 판단된다. 그러나 2001년 부동산 투자회사법과 2003년 간접투자 자산운용법 도입으로 <표 1> 기술통계량에서 보이는 바와 같이 국내자본들도 간접투자자로 오피스빌딩을 매입하는 경우가 점차 증가되고 있는 것으로 판단된다.

'월 임대료'의 경우 1천 원 증가할 때 직접투자 주체가 매입할 *Odds Ratio*보다 간접투자자로 매입할 *Odds Ratio*가 1.32배 증가하는 것으로 보아 임대료가 높아도 교통이 편리하고 수요가 많은 곳을 선호하기 때문인 것으로 판단된다.

'회사채'는 1% 증가할 때 직접투자 주체가 매입할 *Odds Ratio*보다 간접투자자로 매입할 *Odds Ratio*가 0.66배 감소하는데 이는 '회사채'의 수익률이 증가하면 부동산으로의 수익률이 상대적으로 감소하기 때문으로 판단된다.

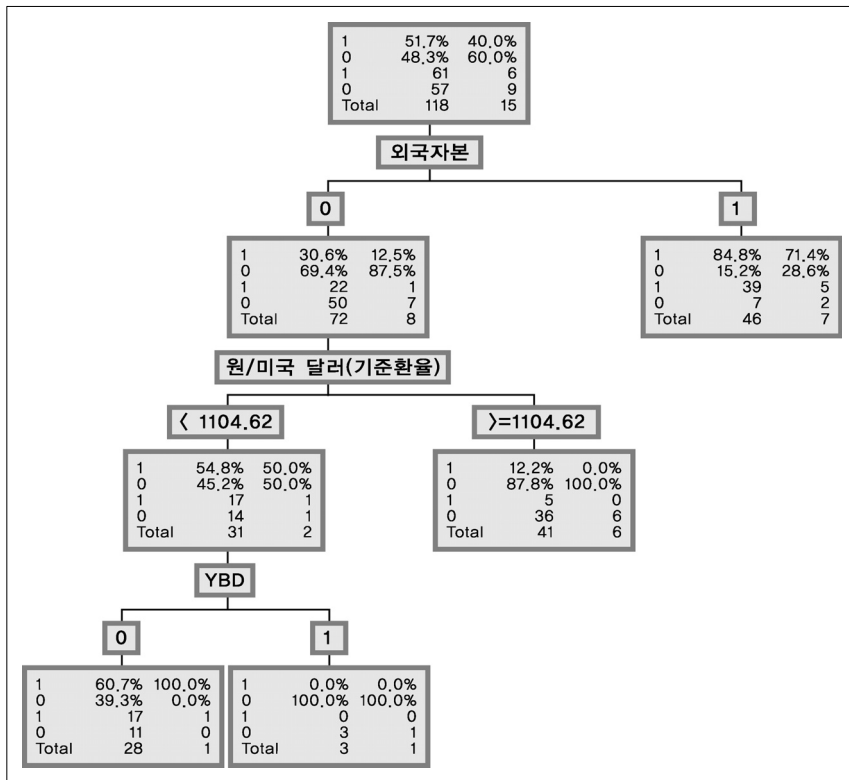
또한 '환율(원화/달러)'이 1원 증가할 때 직접투자 주체가 매입할 *Odds Ratio*보다 간접투자자로 매입할 *Odds Ratio*가 1배로 직접투자자와 차이가 없었다.

위의 결과를 종합해 보면, 간접투자자는 외국자본으로 투자하는 경우 회사채의 수익률이 낮을 때 임대료가 좀 더 높은 서울시 오피스빌딩에 투자하는 것으로 분석되었다.

2) 의사결정나무 분석 결과

매입방식에 영향을 미치는 다양한 독립변수들을 분석한 결과 간접투자자에 대한 의사결정나무 분석은 <그림 6>과 같이 나왔다. '간접투자(1)'는 61건, '직접투자(0)'는 57건으로 총 118건의 데이터가 사용되어 분석되었다. 먼저 최초 가지치기는 '외국자본'의 유무를 기준으로, '외국자본(외국자본=1)' 46건과 '국내자본(외국자본=0)' 72건으로 가지치기를 하였다. '외국자본(외국자본=1)'의 가지를 보면 총 46건 중 39건인 84.8%가 '간접투자(1)'로, 7건인 15.2%가 '직접투자(0)'로 오피스빌딩을 매입하는 것으로 분석되었다.

'국내자본(외국자본=0)'으로 매입한 오피스빌딩을 보면 총 72건 중 50건인 69.4%가 '직접투자(0)'되었으며, 22건인 30.6%가 '간접투자(1)'되고 있는 것으로 분석되었다. '국내자본(외국자본=0)'의 가지는 다시 '기준환율'로 가지를 치고 있다. '기준환율'이 1,100원 미만인 경우 총 31건 중 17건인 54.8%가 '간접투자(1)'로, 14건인 45.2%가 '직접투자(0)'로 매입하였다. 기준환율이 1,100원 미만의 가지는 다시 '여의도 권역' 여부로 가지를 치는데 총 31건 중 '여의도 권역(1)'인 경우 3건이 직접투자(0)되고 있었으며, '여의도 권역이 아닌 권역(0)'은 총 28건 중 17건인 60.7%가 '간접투자(1)'되고,



〈그림 6〉 의사결정나무 분석 결과

11건인 39.3%가 '직접투자자(0)'되고 있는 것으로 분석되었다. 또한 '기준환율'이 1,100원 이상인 경우 총 41건이 투자되고, 이 중 36건인 87.8%가 '직접투자자(0)'로 매입한 것으로 분석되었으며, 이 경우 5건인 12.2%가 '간접투자자(1)'되고 있는 것으로 분석되었다.

위의 결과를 종합해 보면 간접투자자에 가장 많은 영향을 미치는 요인은 외국자본이며, 국내자본으로 간접투자를 하는 경우 기준환율이 1,100원 미만일 때 여의도 권역이 아닌 권역을 선호하는 것으로 판단된다.

3) 신경망 분석 결과

SAS E-MINER 프로그램을 이용하여 학습시켜 학습의 반복으로 오차율을 줄여주고, 오차율의 변동이 거의 없을 때 자동으로 멈추어 결과값을 산출했다. 또한 예측의 정확성을 조금 더 높이기 위해 2개의 은닉층을 가지고 분석하였다. 신경망 분석의 결과 2개의 은닉마디를 가진다는 것을 알 수 있었다.

〈표 4〉는 최종 신경망 모형의 계수를 수치화한 표이며, 식 (3)에 의하여 H_1 을 구한 후, 매입방식(Y)을 산출하였다(b_n =BIAS의 weight, w_{pq} =각 독립변수의 weight, X_p =각 독립변수의 관측값).

신경망 분석의 모수 추정치의 반응확률에 대한 $\hat{P}(Y = \text{간접투자}) = g(-0.11 - 3.60H_{21} + 0.38H_{22})$ (8)
예측식을 도식화하면 식 (8)과 같다.

〈표 4〉 신경망 분석의 가중치표

Parameter		To	weight	To	weight	To	weight
권역	CBD	H_{11}	0.14	H_{12}	-0.39	H_{13}	-0.28
	KBD	H_{11}	0.03	H_{12}	-0.07	H_{13}	-0.26
	YBD	H_{11}	-0.78	H_{12}	-0.23	H_{13}	-0.54
지하철 접근성		H_{11}	-0.66	H_{12}	-0.12	H_{13}	-0.31
연면적		H_{11}	0.33	H_{12}	-0.56	H_{13}	0.36
거래면적		H_{11}	-0.21	H_{12}	0.69	H_{13}	0.01
노후년도		H_{11}	-0.43	H_{12}	-0.01	H_{13}	0.24
외국자본		H_{11}	0.50	H_{12}	-0.48	H_{13}	-0.41
월 임대료		H_{11}	0.33	H_{12}	-1.11	H_{13}	-0.01
월 관리비		H_{11}	-1.01	H_{12}	-0.25	H_{13}	0.52
㎡당 금액		H_{11}	0.06	H_{12}	0.41	H_{13}	0.22
환율		H_{11}	-0.53	H_{12}	-0.58	H_{13}	-0.28
회사채		H_{11}	-1.01	H_{12}	0.59	H_{13}	1.11
BIAS		H_{11}	-1.04	H_{12}	-1.11	H_{13}	0.48
H_{11}		H_{21}	-0.71	H_{22}	-0.88		
H_{12}		H_{21}	0.83	H_{22}	-0.21		
H_{13}		H_{21}	0.55	H_{22}	1.03		
BIAS		H_{21}	-0.16	H_{22}	-1.55		
H_{21}		간접투자	-3.60				
H_{22}		간접투자	0.38				
BIAS		간접투자	-0.11				

〈표 3〉의 로지스틱 회귀분석의 회귀계수 추정치는 식 (9)와 같다.

$$\begin{aligned}
 H = & 8.82 + 0.40 \times CBD + 0.28 \times KBD + 0.4 \\
 & 4 \times YBD + 0.01 \times \text{지하철접근성} + 0.00 \\
 & \times \text{연면적} - 0.00 \times \text{거래면적} - 0.02 \times \text{노} \\
 & \text{후년도} + 1.71 \times \text{외국자본} + 0.27 \times \text{월 임} \\
 & \text{대료} - 0.43 \times \text{월 관리비} - 0.00 \times m^2\text{당} \\
 & \text{금액} - 0.00 \times \text{환율} - 0.42 \times \text{회사채}
 \end{aligned} \quad (9)$$

위의 신경망 분석과 로지스틱 회귀분석의 예측

식에서 보이는 바와 같이 신경망은 매우 유연하기는 하지만 결과를 해석하는 것이 매우 어렵다. 즉, 로지스틱 회귀분석이 제공하는 것과 같은 계수들에 대한 간편한 해석이 불가능하여 어떤 입력변수가 중요한지 또는 그것들이 어떻게 상호작용하는지를 결정하기 어렵다.

데이터 마이닝에서 해석의 용이함이 언제나 예측모형의 중요한 특성이 되는 것은 아니다. 더 많은 해석적 용이함을 가지고 있으면서도 덜 효과적인 모형보다는 매우 정확한 예측을 생산해 내는 신경망이 더 선호되는 경우가 많기 때문이다. 그러므로 본 연구에서는 배화수 외(2008)의 설명에 따라 예측력을 비교, 평가하는 데 신경망 분석을 사용하였다.

4) 모형의 평가

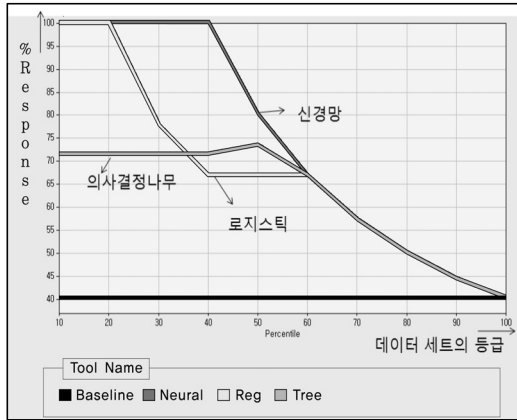
일반적으로 검증용 데이터를 통해 민감도와 특이도, 정확도를 분석한다. 이러한 이유로 본 연구에서도 검증용 데이터를 통해 로지스틱 회귀분석, 의사결정나무 분석, 신경망 분석을 비교해 보았다.

(1) Lift 그래프 통한 모형평가

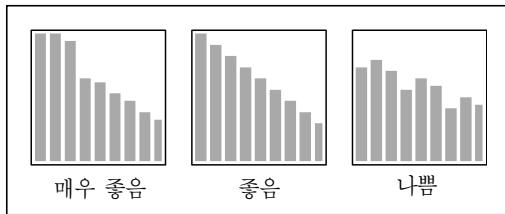
세 모형의 Cumulative Lift 그래프를 보면 <그림 7>과 같다. Cumulative Lift 그래프에서 수직축은 %Response값으로 확률의 크기에 따라 데이터 세트를 내림차순으로 정렬한 다음 각 데이터에 등급을 매겨 각 등급에서 간접투자의 비율을 나타낸다. 수평축은 데이터 세트의 등급을 나타내며 이는 간접투자에 속하는 데이터들이 각 등급에 얼마나 분포하고 있는지를 나타낸다. Cumulative Lift 그래프는 기준선 반응률(그래프에서 Baseline인 %Response값 40%)에 비해 각 등급에서의 반응률이 얼마나 높은지를 나타낸다.

<그림 7> Cumulative Lift 그래프를 보면 로지스틱 회귀분석 모형이 상위 20%에서는 %Response값이 77.8%로 의사결정나무 분석 모형 71.4%보다는 높았지만 신경망 분석 모형 100%보다는 낮았으며 이 지점에서 로지스틱 회귀분석 모형이 급격히 감소하는 것으로 나타났다. 그리고 상위 40%에서는 의사결정나무 분석 모형 71.4%, 로지스틱 회귀분석 모형 66.7%로 두 모형에서는 별다른 차이가 없었으나 신경망 분석 모형이 100%로 높았으며, 이 지점 이후부터는 신경망 분석 모형이 급격히 감소하고 있다. 또한 전체 %Response값이 40%임을 고려할 때 %Response값이 상위 20% 내에 대해서는 로지스틱 회귀분석과 신경망 분석의 경우 약 2.5배의 효율을 얻을 수 있음을 의미한다.

세 모형의 비교를 <그림 8> Cumulative Lift 그래프의 평가와 비교하였을 때 로지스틱 회귀분석은 좋은 모형이고, 신경망 분석의 예측모형이 매우 좋은 모형임을 알 수 있다.

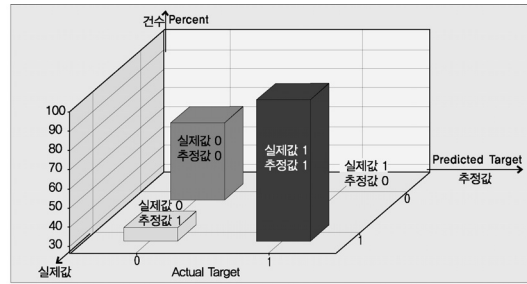


〈그림 7〉 Cumulative Lift 그래프

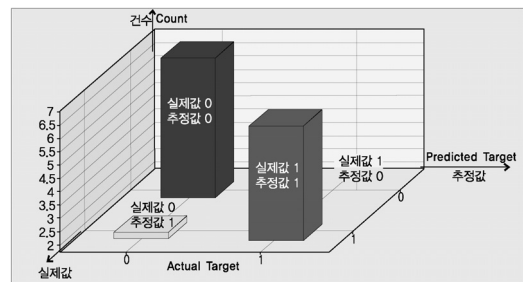


〈그림 8〉 Cumulative Lift 그래프의 평가

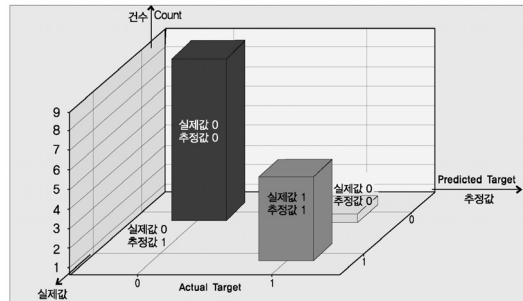
(2) Threshold-based 그래프를 통한 모형평가
로지스틱 회귀분석의 Threshold-based 그래프는 〈그림 9〉와 같고, 의사결정나무 분석의 Threshold-based 그래프는 〈그림 10〉과 같다. 신경망 분석의 Threshold-based 그래프는 〈그림 11〉과 같다.



〈그림 9〉 로지스틱 회귀분석의 Threshold-based 그래프



〈그림 10〉 의사결정나무 분석의 Threshold-based 그래프



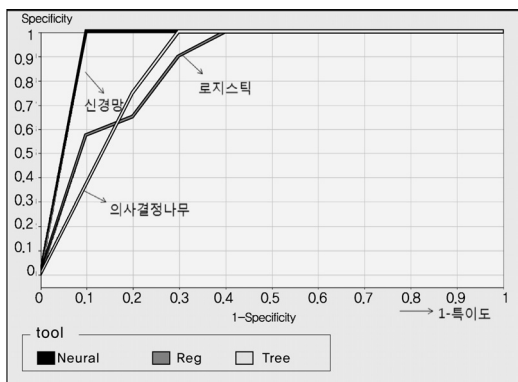
〈그림 11〉 신경망 분석의 Threshold-based 그래프

그래프를 보면 예측값과 실제값의 좌표에서 0은 직접투자, 1은 간접투자를 의미한다. 수직축의 값은 간접투자, 직접투자의 예측값과 실제값의 교차한 결과의 값을 의미한다. 그래프를 보면 예측값 1(간접투자)과 실제값 0(직접투자)일 때와 실제값 0(직접투자)과 예측값 1(간접투자)일 때는 오류값으로 세 모형의 오차가 비슷하다는 것을 알 수 있다. 즉, 세 모형의 정확도가 거의 비슷하지만 로지스틱 회귀분석의 오류값이 약간 더 높은 것으

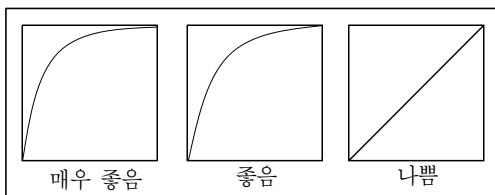
로 보아 로지스틱 회귀분석의 정확도가 약간 떨어진다는 것을 알 수 있다.

(3) ROC 그래프를 통한 모형평가

세 모형의 ROC 그래프를 보면 <그림 12>와 같다. ROC 그래프에서 수직축은 민감도를 수평축은 1-특이도를 나타내며 그 연결선이 대각선의 ROC 곡선이다. 위의 <그림 12>를 보면 <그림 13> ROC 그래프의 평가에서 보이는 바와 같이 세 모형 모두 ROC가 좋은 모형이지만 로지스틱 회귀분석이나 의사결정나무 분석의 모형보다 신경망 분석의 예측 모형이 더 좋은 성능을 가지고 있다고 할 수 있다.



<그림 12> ROC 그래프



<그림 13> ROC 그래프의 평가

(4) 정분류표를 통한 모형평가

식 (5), 식 (6), 식 (7)의 결과로 계산된 정분류표를 통해 세 모형을 평가해 보면 다음과 같다.

그래프의 예측값 1(간접투자)과 실제값 0(직접투자)일 때의 값은 <표 5>의 정분류표에서 수치로 나타내고 있는데 실제값 0(직접투자)과 예측값 1(간접투자)일 때 로지스틱 회귀분석에서는 33%(3개), 의사결정나무 분석에서는 22%(2개)이고, 실제값 1(간접투자)과 예측값 0(직접투자)일 때 신경망 분석에서는 17%(1개)로 로지스틱 회귀분석의 오차가 좀 더 높았다.

위의 결과를 통해 정확도는 로지스틱 회귀분석 80%, 의사결정나무 분석 87%, 신경망 분석은 93%로 신경망 분석이 높았다. 특이도의 경우도 로지스틱 회귀분석이 67%, 의사결정나무 분석 78%, 신경망 분석이 100%로 신경망 분석이 높았다. 그러나 민감도에서는 로지스틱 회귀분석과 의사결정나무 분석은 100%이지만 신경망 분석은 83%로 신경망 분석이 낮았다.

정분류표를 통해 결과를 종합해 보면 특이도, 정확도의 분석 결과는 신경망 분석의 예측력이 높았으며, 민감도의 분석결과는 신경망 분석이 예측력이 낮았다.

위의 예측력의 평가를 통하여 신경망 분석이 다른 분석기법보다 오차가 적고 예측력이 가장 높은 모형이라고 할 수 있다. 그러나 신경망 분석은 예측력이 좋지만 해석이 어려웠고, 로지스틱 회귀분석의 경우 선형성과 정규성의 가정과 잔차의 가정에 의한 정확한 모형을 구축하기 힘들었으며 많은 시간과 노력을 할애하지만, 해석은 편리하였다. 의사결정나무 분석의 경우 모형을 구축하는 시간이 오래 걸리지 않았고, 모형이 도표화되어 모형을 이해하기 편리하였다.

로지스틱 회귀분석이나 의사결정나무 분석은 신경망 분석보다 예측력이 낮았지만 중요한 분석의 의의를 포함하고 있다. 로지스틱 회귀분석과 의사결정나무 분석의 결과 간접투자를 하는 경우

〈표 5〉 세 모형의 정분류표

구분			예측값		합계	예측력	
			직접투자자(0)	간접투자자(1)			
로지스틱 회귀분석	실제값	직접투자자(0)	6(67%)	3(33%)	9(60%)	특이도	67%
		간접투자자(1)	0(27%)	6(100%)	6(40%)	민감도	100%
	합계		6(40%)	9(60%)	15(100%)	정확도	80%
의사결정 나무분석	실제값	직접투자자(0)	7(78%)	2(22%)	9(60%)	특이도	78%
		간접투자자(1)	0(0%)	6(100%)	6(40%)	민감도	100%
	합계		7(47%)	8(53%)	15(100%)	정확도	87%
신경망 분석	실제값	직접투자자(0)	9(100%)	0(0%)	9(60%)	특이도	100%
		간접투자자(1)	1(17%)	5(83%)	6(40%)	민감도	83%
	합계		10(67%)	5(33%)	15(100%)	정확도	93%

회사채의 수익률이 낮을 때, 외국자본으로 임대료가 좀 더 높은 서울시 오피스빌딩에 투자하고, 국내자본으로 투자하는 경우 기준환율이 1,100원 미만일 때 여의도 권역이 아닌 권역을 선호하는 것을 알 수 있었다.

IV. 연구의 한계 및 향후 과제

본 연구는 1998년부터 2009년 1/4분기까지 16,500㎡ 이상의 서울시 오피스빌딩 거래사례를 통해 간접투자자로 투자하는 경우 어떠한 요인들을 고려하는지를 분석하기 위해 로지스틱 회귀분석, 의사결정나무 분석 그리고 신경망 분석을 사용하였고, 그 세 분석모형의 예측력을 비교, 평가하는데 그 목적이 있다.

로지스틱 회귀분석을 통하여 간접투자하는 경우 회사채의 수익률이 낮을 때 외국자본으로 임대료가 좀 더 높은 서울시 오피스빌딩에 투자하는 것으로 분석되었다.

로지스틱 회귀분석에서 얻어진 각 변수들의 계수의 영향력 해석을 위해 의사결정나무 분석을 사용하였다. 이를 통해 간접투자 시 고려되는 요인

들 중 어떠한 요인들이 영향을 주는지에 대해 분류된 요인 분석을 하였다. 그 결과 간접 투자방식에 가장 많은 영향을 미치는 요인은 외국자본이며, 국내자본으로 투자하는 경우 기준환율이 1,100원 미만일 때 여의도 권역이 아닌 권역을 선호함을 알 수 있었다.

두 결과를 종합해 보면 간접투자자는 외국자본으로 투자하는 경우 회사채의 수익률이 낮을 때 임대료가 좀 더 높은 서울시 오피스빌딩에 투자하고, 국내자본으로 투자하는 경우 기준환율이 1,100원 미만일 때 여의도 권역이 아닌 다른 권역을 선호함을 알 수 있었다.

세 모형의 예측력을 토대로 Cumulative Lift 그래프가 그려졌는데 로지스틱 회귀분석 모형이 상위 20%에서는 %Response값이 77.8%로 의사결정나무 분석 모형 71.4%보다는 높았지만 신경망 분석 모형 100%보다는 낮았으며 로지스틱 회귀분석 모형이 급격히 감소하는 것으로 나타났다. 그리고 상위 40%에서는 의사결정나무 분석 모형 71.4%, 로지스틱 회귀분석 모형 66.7%로 별다른 차이가 없었으나 신경망 분석 모형 100%로 높았으며, 신경망 분석 모형이 급격히 감소하고 있다. 전

체 %Response값이 40%임을 고려할 때 %Response 이 상위 20% 내에 대해서 로지스틱 회귀분석과 신경망 분석의 경우 약 2.5배의 효율을 얻을 수 있음을 의미한다.

또한 모형 모두 ROC가 좋은 모형이지만 로지스틱 회귀분석이나 의사결정나무 분석의 모형보다 신경망 분석의 예측모형이 더 좋은 성능을 가지고 있다고 분석되었다.

세 가지의 분석 모형에 대한 예측력이 높은 모형을 선별하기 위해 정분류표에 의해 정확도, 민감도, 특이도를 측정하였다. 그 결과 정확도는 로지스틱 회귀분석이 80%, 의사결정나무 분석 87%, 신경망 분석 93%로 신경망 분석이 제일 높았다. 특이도의 경우도 로지스틱 회귀분석이 67%, 의사결정나무 분석 78%, 신경망 분석 100%로 신경망 분석이 제일 높았다. 그러나 민감도의 경우 로지스틱 회귀분석과 의사결정나무 분석은 100%이지만 신경망 분석은 83%로 제일 낮았다.

종합적으로 Cumulative Lift 그래프, ROC 그래프에서 신경망 분석이 매우 좋은 모형으로 분석되었으며, 정확도, 특이도도 신경망 분석이 높았다. 결과적으로 예측력에서는 신경망 분석이 가장 좋은 모형으로 분석되었다.

간접투자에 관한 연구는 대부분이 외국의 간접투자제도와 비교·분석이나 우리나라 간접투자제도의 문제점 및 활성화 방안을 연구하는 데 그치고 있다. 왜냐하면 오피스빌딩의 경우 거래가 격이나 임대료, 영업경비와 같은 자료는 소유자의 영업비밀로 취급되어 공개를 꺼리는 시장분위기와 간접투자된 국내외 자본이 선호하는 오피스빌딩의 유형에 관한 데이터 수집이 어렵기 때문이다. 이런 상황에서 투자방식에 따른 양질의 데이터로 실증분석을 하였다는 점에서 본 연구의 의의

가 있다.

그러나 본 연구는 이항 로지스틱 회귀분석을 사용하여 간접투자와 직접투자와의 독립변수의 차이만을 비교하였다. ABS, REITs, REF를 간접투자로 분석하기보다는 각각을 종속변수로 넣어서 분석함으로써 각각의 특징을 분석하지 못한 아쉬움이 남는다. 이는 간접투자된 오피스빌딩 데이터의 양이 많지 않았기 때문이다. 또한 외국자본이 투자하는 경우 자회사 주식인수방식도 간접투자로 볼 수 있지만 본 연구에서는 공시되어 확인된 내용에 한해서만 간접투자로 하여 분류하였기 때문에 이를 고려하지 못하였다. 이에 대한 연구의 한계를 향후 과제로 남긴다.

참고문헌

- 강현철 · 한상태 · 최종후 · 이성건 · 김은석 · 엄익현 · 김미경, 2009, 「고객관계관리(CRM)를 위한 데이터 마이닝 방법론」, 자유아카데미.
- 강혜진 · 오동훈, 2009, “부동산 간접투자 오피스빌딩의 특징 및 수익률과의 상관관계 분석”, 『주택연구』, 17(1): 95~118.
- 김상용, 2009, “국내외 자본의 서울 오피스빌딩 투자행태에 관한 연구”, 건국대학교 부동산대학원 석사학위논문.
- 김선주 · 이춘섭, 2005, “부동산 펀드의 자산 운용에 관한 연구: 오피스 실증자료를 중심으로”, 건축대학교 석사학위논문.
- 배화수, 2008, 「SAS Enterprise Miner를 이용한 데이터마이닝(개정판)」, 서울: 교우사.
- (주)신영에셋, 2009, 「오피스 시장동향보고서」.
- 이상경 · 이인철, 2005, “외국자본의 서울 오피스빌딩 투자행태 연구”, 『국토계획』, 40(5): 177~188.
- 이준용, 2008, “외국자본의 서울시 오피스빌딩 투자에 대한 입지적 특성 연구”, 건국대학교 대학원 석사학위논문.
- 이준용 · 최미화 · 이상엽, 2008, “데이터 마이닝 적용을 통한 아파트 가격 예측에 관한 연구”, 『국토계획』,

42(4): 135~148.

R2Korea, 2009, *Office Market Report*.

정광모, 2003, 「로지스틱 회귀와 응용」, 자유아카데미.

홍아름 · 한제선 · 유선중, 2009, “서울시 오피스빌딩 간접

투자요인에 관한 연구”, 『주택연구』, 17(4): 55~76.

홍종선 · 박용석, 2005, 「통계적 신용평가모형의 효율성 비

교」, 93~107, 성균관대학교 응용통계연구소.

원 고 접 수 일 : 2010년 1월 12일

1차심사완료일 : 2010년 3월 5일

최종원고채택일 : 2010년 4월 30일