



# Journal of Real Estate Analysis

<http://www.kab.re.kr>

November 2017, Vol.3, No.2, pp.39-56

## 데이터마이닝을 이용한 부동산펀드 위험 특성 분석

### Analytics Using Data Mining for Forecasting the Risk Characteristics of Real Estate Funds

김경민\*  
Kim, Gyung Min

#### Abstract

This study was carried out with the intent to help institutional investors and the staff of investment companies to select investment goods on the basis of the results of forecasting and analyzing the risk characteristics of domestic and foreign real estate funds (REFs) through classification analysis, thereby minimizing risk and maximizing profitability. The data used for the valuation analysis and classification analysis of REF risk consisted of data on 402 asset management companies' performance in domestic and overseas REF investment products. Data mining was used to classify and analyze the risk characteristics of REFs. As a result of the predictive analysis of investment values based on the classification analysis of data mining, the classification accuracy of the predictive analysis was found to be 97.76%, showing a high classification rate. Therefore, it is expected that the results of analysis using the decision tree model (C5.0) will offer standards for the development and sale of domestic and overseas REF products to the persons in charge of product planning and asset management in asset management companies and product sale companies.

Key words : Domestic and overseas Real Estate Fund, Severity Rating, Risk prediction, Classification Analysis, Sorting Rate

\* 강남대학교 부동산건설학부 외래교수 / 한국데이터진흥원 강사 / (주)골드앤코리아리얼티 대표  
(kkmin0390@naver.com)

## 1. 서론

### 1. 연구의 배경 및 목적

저금리·저성장애 따라 국내외 부동산 간접 투자 상품으로서의 부동산펀드에 대한 관심이 높아지고 있다. 국내외 증시 흐름이 불안한 가운데 부동산 간접투자 상품인 부동산펀드는 기관 뿐 아니라 개인 투자자들에게도 투자 포트폴리오 구성에 있어 안정적인 수익이 기대되고 있기 때문이다.

주식과 채권 모두 투자전망이 불확실한 상황에서 대체투자수단으로 부동산간접투자가 연기금 등 기관투자자들을 중심으로 투자가 집중되고 있다.<sup>1)</sup>

부동산간접투자의 증대와 자본 시장에서의 강력한 영향력으로 인한 시장의 공개요구로 언론의 불균형 또한 심화되고 있다.<sup>2)</sup> 이러한 언론의 부동산펀드에 대한 판단은 국내 자본시장 및 부동산시장에 있어 유동화 및 파생을 통한 부동산 간접투자 증대에 있어 위험성 판단에 장애요인으로 작용할 수도 있는 경향을 내포하고 있다.

국내에 판매되는 부동산펀드 판매상품 중 해외부동산펀드 상품의 비중이 급증하고 있다. 해외 부동산 펀드 대부분은 투자위험등급이 1~2등

급(매우높은 위험~높은 위험)이다. 환율과 금리 변동에 따라 마이너스 수익을 낼 수도 있다.<sup>3)</sup>

일반적으로 투자자들은 고수익의 성취에 대한 관심보다는 자신의 돈을 잃는 것에 더 민감한 반응을 나타내는 것처럼 보인다.<sup>4)</sup>

국내의 부동산펀드는 수익에 비하여 위험성 또한 높다. 부동산 펀드 특성상 환매가 어려워 장기간 자금이 묶일 가능성은 감안해야 한다. 매각시점의 악재로 부동산가격이 떨어지며 오랜 기간 동안 투자금이 동결되는 피해를 볼 수 있다. 또한 부동산펀드 상품들이 투자하는 부동산이 금리 인상 등 변동 요인으로 큰 손실을 볼 우려가 있다.

본 연구는 국내외 부동산펀드 상품군의 위험 특성을 분석하여 가장 높은 위험등급인 1등급 위험(risk)의 분류분석 결과를 바탕으로 투자 기관 및 개인들로 하여금 투자상품 선별에 조력을 주어 위험을 최소화하여 손실을 최소화시켜 수익성을 극대화 시키고자 하는데 연구의 주된 목적을 두었다.

### 2. 연구의 범위 및 방법

본 연구는 분류분석을 통한 부동산펀드 위험 특성을 예측하기 위하여 국내외 투자상품 402개

1) 박덕배, 2015, “부동산 간접투자시장의 특징과 시사점”, VIP REPORT(15-7호), 현대경제연구원, p.7.

2) Scott R. Muldavin, 2001.8., The New World of Real Estate Finance, Real Estate Finance, The Maldavin Company, Inc., p.75.

3) 실제로 2012년 설정된 '미래에셋맵스프런티어브라질' 상품- 5년 수익률이 -53%에 달한다. 브라질 헤알화 가치가 급락하면서 수익률이 급감했다.

4) 노희진, 2009, 『펀드경영론』, 박영사, p.195.

의 펀드 판매사들의 실적데이터를 데이터마트로 구성하였다. 부동산펀드 상품들의 예측모형 판단의 기간은 최근 3년간의 기준을 2017.11월 말(2014년 12월~2017년 11월)로 하여 1년 단위로 하였다. 이를 통해 투자가치 판단을 위해 정제된 데이터마트를 구성하여 데이터마이닝을 활용한 기계학습 예측분석을 실시한다. 이러한 데이터마이닝을 활용해 예측력을 확보하여 부동산펀드의 투자특성을 분석한다.

국내 자산운용사가 판매사를 통하여 판매한 부동산펀드 상품 위험 판단을 통한 투자성과 예측을 위하여 데이터마이닝을 이용한 기계학습 분석으로 서포트벡터머신(SVM), 로지스틱회귀분석(LR), 의사결정나무(C5.0), 인공신경망(ANN)을 실시하였다.

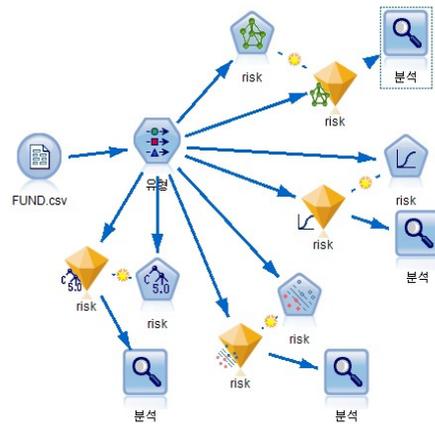
선행연구 및 기초이론검토를 통하여 부동산펀드 투자 위험 관련 선행연구를 분석하여 본 연구와의 차별성을 부각시킨다. 또한 자산운용사의 부동산펀드 투자 현황 및 투자수익률을 확인하여 대체투자로서 부동산펀드의 투자성과를 평가한다.

실증분석은 국내 자산운용사 및 판매사의 국내 부동산펀드 투자 실적을 원데이터로 하여 데이터마이닝 분류분석을 이용한 펀드투자 위험 특성 예측 분석을 실시한다.<sup>5)</sup> 이를 위해 분류의 정확도 분석을 실시하고 정분류율을 통한 예측

력 확보를 바탕으로 종속변수인 투자성과에 중요한 영향을 미치는 독립변수를 검증한다.

분석 데이터 마트 구성을 위한 원시데이터는 금융투자협회 데이터를 참조하였다. 국내외 부동산펀드의 위험 특성 비교분석을 위하여 금융투자협회 국내외 부동산펀드 판매 데이터<sup>6)</sup>를 활용하였다.

데이터마이닝 분류분석 위험예측에 활용한 실증분석 프로그램은 SPSS22.0, SPSS Modeler를 활용하였다.



〈그림 1〉 데이터마이닝 분석 모형

실증분석 연구모형으로는 인공신경망(ANN), 서포트벡터머신(SVM), 로지스틱회귀분석(LR), 의사결정나무(C5.0) 모형을 활용하여 분류의 정확

5) 데이터마이닝을 명확히 정의하기는 어렵다. 대용량의 자료로부터 정보의 요약과 미래에 대한 예측을 그 목표로 하며 자료에 존재하는 관계, 패턴, 규칙 등을 탐색하고 이를 통계적으로 모형화 함으로써 이전에는 알려지지 않은 유용한 지식을 추출하는 일련의 과정으로 정의할 수 있다(박창이·김용대·김진석·송중우·최호식, 2015, 『R을 이용한 데이터 마이닝』, 교우사, p.3.).

6) 금융투자협회(<http://dis.kofia.or.kr/websquare/index>).

도와 예측력을 비교분석하였다. 이에 대한 모형 도식은 <그림 1>과 같다.

## II. 선행연구 및 분석방법

### 1. 선행연구

국내의 선행연구는 주로 리츠 또는 부동산펀드의 다른 자산과 함께 포트폴리오를 구성할 경우 리스크는 낮추고 수익률은 안정화시키는 포트폴리오 효과를 논증하거나 거시경제변수가 리츠 수익률 또는 부동산펀드 수익률에 어떠한 영향을 끼치는지를 연구한 논문이 주류를 이루었다.<sup>7)</sup> 부동산간접투자상품에 대한 선행연구로서의 부동산펀드에 관한 선행연구는 부동산펀드 성과요인, 수익률 영향 요인에 관한 연구가 주류를 이루고 있다. 선행연구를 종합하면 부동산펀드의 성과는 거시경제요인과 투자대상 부동산의 특성을 비롯하여 펀드운영자의 능력 등 제반 영향요인들이 복합적으로 작용한 결과라고 볼 수 있다.

김종원(2013)<sup>8)</sup>, 김은주(2010)<sup>9)</sup>를 비롯한 다수

연구자들이 부동산펀드와 거시경제지표와의 관련성을 연구하였다. 이경순(2010)<sup>10)</sup>은 2009년 10월 기준 리츠와 부동산펀드 가운데 임대형 자산(오피스, 상업용)에 투자한다. 설립 후 6개월이 지난 상품 중 자료를 구할 수 있는 공모형 리츠 10개, 사모형 리츠 14개, 부동산펀드 7개를 대상으로 배당수익률 결정요인을 분석하였다. 보수운용비용, 운용기간, 부동산 개수, 총자산의 크기가 영향을 미친다고 하였다. 김재훈(2008)<sup>11)</sup>은 리츠의 자기자본수익률의 지표인 배당률과 유의한 관계가 있는 독립변수로 LTV, 연간이자율, 자산관리회사, 부동산 종류, 리츠의 종류임을 실증하였다. 장용삼 외 2인(2011)<sup>12)</sup>은 리츠와 타 금융자산의 수익위험 분석을 통해 리츠의 특성을 확인하였다.

해외 선행연구에서 David Enke(2005)<sup>13)</sup>는 데이터 마이닝을 위한 기계학습에 사용되는 정보 획득 방법을 소개하였다. 다수의 금융 및 경제변수의 예측 관계를 평가하였다. 수준 추정 및 분류를 위한 신경망 모델이 효과적인 미래 가치 예측을 위해 조사되었다.

K.C.Tseng(2010)<sup>14)</sup>은 '데이터마이닝을 활용한 금융주식 예측' 연구에서 주가의 예측을 위한

7) 남연우·고석찬, 2015, 「부동산펀드 성과요인 분석」, 『부동산학보』, 제62집, p.83.

8) 김종원·최민섭, 2013, 「부동산 간접투자 수익률에 영향을 미치는 요인에 관한 연구-부동산펀드를 중심으로」, 『부동산학연구』, 제19집, pp.143~160.

9) 김은주·고성수, 2010, 「경제환경에 따른 부동산펀드의 수익률 형태」, 『부동산연구』, 제19집, pp.51~73.

10) 이경순, 2010, 「임대형부동산 간접투자상품의 배당수익률 결정요인분석」, 건국대학교 석사학위논문, pp.1~185.

11) 김재훈, 2008, 「리츠의 자기자본수익률에 영향을 미치는 요인에 관한 실증 연구」, 연세대학교 석사학위논문, pp.1~66.

12) 장용삼·김진호·김행중, 2008, 「부동산투자신탁의 상관성에 관한 연구」, 『한국지적학회지』, 제24권 제1호, pp.95~108.

13) David, Enke·Suraphan Thawomwong, 2005, "The use of data Mining and neural networks for forecasting stock market returns", *Expert Systems with Application*, Vol.29., pp.927~940.

14) K.C.Tseng, 2010, "Forecasting Financial Stocks Using Data Mining", California State University Fresno, Forecasting, pp.1~24.

예측의 정확성을 증가시키기 위해 인공신경망(ANN)과 OLS모형을 비교하였다.

본 연구의 기존 선행연구와의 차별성은 기존 부동산펀드 관련 선행연구들이 부동산펀드의 성과요인 및 수익률 영향요인 분석이 위주였으나 자산운용사 및 판매사를 통하여 실무에서 투자된 부동산펀드 상품 실적 실무 데이터를 바탕으로 데이터마이닝 분류분석 기법을 활용한 위험 특성 예측에 중점을 둔 연구라는 점이다.

## 2. 이론적 기초

### 1) 부동산펀드의 위험성

부동산투자에서의 위험이란 부동산투자를 통해 장래 발생하는 수익의 변동성을 말한다.<sup>15)</sup> 부동산펀드 투자 등 투자의 경우에 어떤 투자로부터 미래에 얻을 수 있는 결과를 확실하게 알지 못하는 상황을 불확실성(uncertainty) 또는 위험(risk)이 있는 상황이라 한다.<sup>16)</sup>

부동산펀드(real estate fund)는 다수 투자자로부터 자금을 모아서 공동기금을 조성하여 전문적인 투자기관에 맡겨 부동산, 부동산과 관련한 대출 및 유가증권 등에 투자하여 운용성과에 따라 수익을 분배하는 투자신탁(business trust), 투자회사(corporation) 또는 간접투자상품을 말

한다.<sup>17)</sup> 부동산펀드는 투자수익률이 고정되어 있지 않고 불확실하며 원금손실을 가져올 수 있는 위험성을 가지고 있다.

펀드는 1~5등급으로 투자위험 수준이 나뉘는데 보통 1등급이 가장 기대수익률이 높으나 원금손실 위험도 가장 큰 상품이다. 위험등급별 펀드유형은 1등급~5등급으로 구분되어 있다. 위험등급 1등급펀드에는 1년 만에 필요한 돈을 투자해서는 절대 안된다. 부동산펀드는 일반적으로 3등급 중간 위험을 구성한다. 가장 위험도가 높은 위험 1등급의 기준은 변동성 25%이상이다.<sup>18)</sup>

증권화된 부동산상품에 대한 간접투자는 부동산을 소유하지 않으면서 투자하는 형태로서 부동산 직접투자에서 발생할 수 있는 각종 위험요소를 크게 감소시킬 수 있다.<sup>19)</sup> 부동산 증권화 상품은 부동산자체가 가지고 있는 위험을 금융자체의 위험으로 바꾸어 상품화하고 있다. 부동산펀드는 실물자산이 금융자산화하면서 부동산고유의 리스크<sup>20)</sup>와는 다른 새로운 위험이 발생하게 되었다.

부동산펀드는 그 대상이 되는 프로젝트 사업성의 불확실성에서 오는 위험을 피할 수 없다. 이러한 불확실성이야말로 높은 수익률을 제공할 수 있는 기회요인이기도 하다.

15) 박원석, 2013, 『부동산투자론』, 양현사, p.18.

16) 박정식·박종원, 2008, 『현대투자론』, 다산출판사, p.114.

17) 손재영 외, 2009, 『한국의 부동산금융』, 건국대학교 출판부, p.190.

18) 금융감독원, 2015.12.28., 『보도자료』, pp.1~6.

19) 김형석 외, 2005, 『부동산펀드』, 서울 : 매일경제신문사, pp.141~142.

20) 펀드위험이란 펀드수익률의 변화이다.

## 2) 부동산펀드의 유형 및 가치평가

리츠의 주요한 3가지 유형으로는 지분형(Equity Trusts), 저당대출형(Mortgage Trusts), 혼합형(Hybrid Trusts)이 있다. 지분형이 보유하고 있는 자산과 저당대출형이 보유하고 있는 자산의 차이는 명확하다. 지분형은 부동산에 대한 소유권 지분을 취득하는 반면, 저당대출형은 부동산저당대출채권을 매수하여 부동산 지분 소유자보다 선순위의 저당채권자가 된다. 두 가지 형태의 펀드의 장점을 결합하는 더욱 복잡한 투자 형태가 발전하였는데, 이를 혼합형 펀드라고 한다.<sup>21)</sup>

저당대출형에 있어서는 향후 일반적인 대출형의 경우 보다 수익 참여도가 높은 메자닌금융이 활성화될 것으로 본다.

펀드수익률은 운영성과를 나타내는 지표로 일정기간 동안 기준가격이 얼마나 증감했는지를 나타낸다.<sup>22)</sup> 부동산펀드가 공시하는 자료들은 다양한 정보를 제공한다. 이러한 자료들에는 부동산펀드들이 소유하고 있는 부동산가치를 정확하게 제공해 주지 않는다. 이와 관련된 합리적인 해결책은 일반인이 생각하는 것보다 어려운 작업이다.<sup>23)</sup>

펀드평가 모형으로는 첫째, 순자산가치모형이 있다. 펀드 기간수익률이란 <(종료일 수정기준가 - 시작일 수정기준가)/시작일 수정기준가

\*100(%)>로 계산 한다. 리츠 및 부동산펀드 모두 주로 개인들이 직접 투자하는 주택보다 오피스 등 상업용 부동산에 투자한다. 둘째, 소득승수모형(Income(FFO) Multiple)이 있다. 셋째, 고든 모형(Gorden Dividened Discount Model)이 있다. 이 모델에서 주식의 가치는 차기배당금(D1)을 요구수익률(K)과 배당성장율(G)의 차로 나눈 값이다.

펀드는 한 상품에만 투자하기 보다는 여러 상품에 나누어 투자하는 것이 현명하다. 원금 손실이 발생해도 수익이 회복되기를 기다렸다가 만족할 만한 수익을 올렸을 때 환매할 수 있을 정도로 여유가 있는 자금인지 확인한 후 투자하는 것이 바람직하다.

## 3) 부동산펀드의 현황

공모부동산펀드 시장에 비하여 사모부동산펀드 시장의 성장세가 높은 가운데 기관투자자가 국내외 부동산에 대한 투자를 늘리고 있으며 2010년 이후 해외부동산펀드에 대한 투자가 커지고 있다.<sup>24)</sup>

2004년 도입된 부동산펀드는 수탁고 8,610억 원, 2005년 2조 5,609억 원 2006년 4.1조 원으로 태동하였다. 꾸준히 성장하여 2012년 말 19조 9051억 원, 2016년 10월말 현재 43.6조 원의 펀드수탁고를 기록하고 있다.

21) 김영곤 외 4인, 2014, 『부동산금융과 투자』(제14판), 부연사, pp.745~746.

22) 방경식·장희순, 2016, 『부동산학총론』(제4판), 부연사, p.580.

23) 김영곤 외 4인, 2014, 상계서, p.114.

24) 김규림, 2016, 「최근 국내 부동산펀드시장 현황과 시사점」, 『자본시장 weekly』, 2016-25호, p.3.

〈표 1〉 펀드 유형별 펀드수

(단위 : 건)

년도	주식	채권	부동산	합계
2012	1,118	345	14	3,261
2013	1,100	324	16	3,310
2014	1,161	335	21	3,448
2015	1,216	343	19	3,746
2016	1,147	349	20	3,608

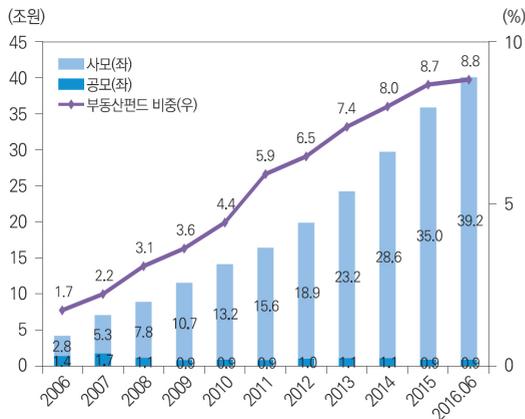
자료 : 금융투자협회(2012~2016).

〈표 2〉 펀드 유형별 순자산

(단위 : 건)

년도	주식	채권	부동산	합계
2012	796,594	129,070	10,484	1,862,931
2013	721,640	124,220	10,935	1,844,114
2014	635,762	152,645	11,254	1,981,465
2015	636,320	189,715	8,671	2,137,869
2016	561,257	239,368	12,742	2,122,156

자료 : 금융투자협회(2012~2016).



〈그림 2〉 국내부동산펀드 시장 추이

자료 : 김규림, 2016, 「최근 국내 부동산펀드시장 현황과 시사점」, 『자본시장 weekly』, 2016-25호, p.3.

### III. 분석자료 및 수익률 추정

데이터마이닝은 모형의 타당성 평가를 위하여 데이터를 훈련데이터(training data)와 평가데이터(test data)로 구분한다.<sup>25)</sup> 훈련데이터는 모형을 생성하는데 이용되고, 평가데이터는 최종적인 평가를 수행하는데 사용된다. 본 연구에서는 훈련 데이터 80%, 검증용 데이터 10%, 평가데이터 10%로 나누어 분석을 실시하였다.

본 연구에서는 국내의 부동산펀드의 수익률 예측분석을 위하여 분류모형(Classification Model)을 통하여 예측의 정확도를 확인하였다. 이러한 예측의 정확성을 바탕으로 펀드 판매사별 부동산펀드로 판매된 최근 3년(2014~2017년) 사례(402N, 매우큰위험의 펀드(1)-230N, 매우큰위험이 아닌 펀드(0)-172N)하여 위험(Risk) 특성을 분류 분석한다. 분류모형에 사용된 이론을 살펴보면 다음과 같다.

#### 1. 검증모형 및 분류모형 평가

##### 1) 검증모형

##### (1) 인공지능망 분석

(Artificial Neural Network, ANN)

신경망(Neural Network)은 뇌구조에 영향을 받았다. 신경망모형((Neural Network Model)은 생물학적 신경망의 구조로부터 착안된 학습 알고리즘으로서, 특히 입력 값을 바탕으로 잠재변수를 만들고 잠재변수를 기반으로 출력 값을 예

25) Training Data는 Training Data와 Validation Data로 구분하여 사용되며 일반적인 구분 비율은 7:3이다.

측하는 비선형구조를 가지는 모형이다.<sup>26)</sup> ANN 은 SVM같이 블랙박스 알고리즘이다. 그래서 뉴런의 이동과 데이터 변이의 구조와 메카니즘은 매우 복잡하고 독창적이지 않다.

뉴런의 기능과 활동 구조식은 다음 (식 1)과 같다.

$$y = \varphi \left( \sum_{i=0}^2 w_i \cdot x_i \right) \quad (\text{식 1})$$

신경망에 있어서의 신경은 정보가 통과하는 다층 구조로 하나의 Input layer, 하나 또는 그이상의 Hidden layer, 그리고 Output layer로 되어 있다. 다계층 신경망은 위에서 언급한 바와 같이 전의 계층의 신경으로부터 데이터를 얻는다. 조작된 데이터는 다음의 Hidden layer 혹은 Output layer의 신경에 전달되어졌다.<sup>27)</sup>

## (2) 서포트벡터머신 분석

(Support Vector Machine, SVM)

SVM(Support Vector Machine)은 비선형분류 알고리즘 중 하나로 가장 효율적이고 유연한 방법 중 하나이다. 초월평면에 가장 근접한 인스턴스들을 Support Vector라 부른다. 선형모델의 최대 단점은 클래스 간의 선형경계 밖에 표현하지 못한다는 점이다.<sup>28)</sup>

예측이 정확하고 여러 가지 형태의 자료에 대하여 적용이 쉽지만 확률 추정 값을 요구하는 분야에는 잘 사용되지 않는다.<sup>29)</sup>

SVM은 서포트벡터 분류를 의미한다. 또한 분류문제뿐 아니라 회귀문제도 적용가능하다.

두 개의 클래스를 분리하는 초월 평면을 다음 (식 2)와 같은 방정식으로 표현할 수 있다.

$$x = w_0 + w_1 a_1 + w_2 a_2 \quad (\text{식 2})$$

최대 마진 초월평면은 클래스를 가장 많이 분리하는 평면에 해당하며 다음 (식 3)과 같다.

$$x = b + \sum_{i \in \text{support}} a_i y_i a(i) \cdot a \quad (\text{식 3})$$

## (3) 로지스틱회귀분석

(Logistic Regression, LR)

로지스틱회귀분석(Logistic Regression, LR)이 선형회귀분석과의 가장 큰 차이점은 타겟 가치의 규모이다. 선형회귀분석은 종속변수가 범주형 변수일 때 이를 처리하기에 부적합하다. 그래서 다른 접근방법이 요구되는데 이것이 로지스틱회귀분석이다. 중심아이디어는 종속변수 자체를 예측하는 대신 가능한 결과의 가능성을 예측하는데 회귀모형의 변환을 수행하는 것이다. 이 모형이 바로 “LR 모델”이다.<sup>30)</sup>

26) Begas, 2016, 『BIG DATA, 빅데이터분석전문가 교재』, 한국데이터진흥원, p.350.

27) Tilo Wendler · Sören Gröttrup, 2016, 『Data Mining with SPSS Modeler』, pp.318~321.

28) 이안위튼 · 아이베프랑크 · 마크홀, 2013, 『데이터마닝(기계학습의 이론과 응용)』, 에이콘, pp.318~321.

29) Begas, 2016, 상계서, p.343.

30) Tilo Wendler · Sören Gröttrup, 2016, 전계서, pp.733~736.

로지스틱회귀분석은 어떤 사건이 발생하는지 안하는지를 직접 예측하는 것이 아니라, 그 사건이 발생할 확률을 예측한다. 따라서 종속변수 값은 0 또는 1사이의 값을 갖는다.

로지스틱 구조는 다음 (식 4)와 같다.

$$F(t) = \frac{\exp(t)}{1 + \exp(t)} \quad (\text{식 4})$$

(4) 의사결정나무 분석(Decision Tree, C5.0)

의사결정나무모형은 비선형 회귀분석, 비선형 판별분석의 하나로 분류되는데, Tree모형 그자체로서 분류, 또는 예측과정으로 수행되기도 한다.<sup>31)</sup>

의사결정나무의 분리규칙은 각 마디에서 분리에 사용될 독립변수의 선택과 분리가 이루어질 기준(분리기준)을 의미한다. C5.0은 이익비율 표준을 사용하여 이익 비율이 최대화되는 점에서 데이터의 분할을 선택하는 다지분리 알고리즘이다. 엔트로피, 정보이득 개념을 사용하여 분리기준을 결정한다.

$$E(S) = \sum_{i=1}^n p_i \log_2 p_i \quad (\text{식 5})$$

$p_i$ 는 자료(S)내에서  $i$ 번째 범주가 차지하는 비율을 말하며,  $n$ 은 범주의 개수이다.

정보이득은 엔트로피의 감소량이며, 엔트로피가 가장 낮아지는 속성을 찾는것이 목적이다.

$$Gain(A) = E(S) - \sum_{k=1}^n E(S_k) \frac{|S_k|}{|S|} \quad (\text{식 6})$$

$Gain(A)$  자료는 자료(S)를 독립변수(A)의 값에 따라  $n$ 개로 나누었을 때 얻게 되는 정보이득이다.<sup>32)</sup>

2) 분류모형의 평가

분류 또는 예측 모형의 유용성을 판단하고 상이한 모형들을 서로 비교하여 평가를 하는 것이 필요하다. 분류 모형의 성과 평가는 분류 정확도를 통해서 나타난다. 분류 정확도는 분류 결과의 정확성을 평가하여 최적의 분류 모형을 선택하는 방법이다.

분류모형 평가를 통해 분류모형 특성에 따라 예측모형의 정분류율과 변수의 중요도에 차이가 발생하였다.<sup>33)</sup> 분류모형에 따른 정확도 성능 측정의 차이도 이를 설명한다.<sup>34)</sup> 분류모형별 정확성을 설명실제집단과 모형에 의해 분류한 집단을 비교하여 그 결과를 다음과 같이 <표 3>으로 요약할 수 있다.

31) 이재길, 2016, 『R프로그래밍에 기반한 다변량분석 및 데이터마이닝』, 향소걸음아카데미, p.321.

32) Begas, 2016, 전개서, pp.300~307.

33) 김나영 · 이은경, 2014, 「나무구조의 분류분석에서 변수 중요도에 관한 고찰」, 『응용통계연구』, p.724.

34) 캐빈 머피, 2015, 『머신러닝』, 에이콘, pp.693~694.

〈표 3〉 데이터의 실제집단과 분류된 집단간의 결과 교차표

구분		분류된 집단	
		A <sub>1</sub>	A <sub>2</sub>
실제 집단	A <sub>1</sub>	B <sub>11</sub>	B <sub>12</sub>
	A <sub>2</sub>	B <sub>21</sub>	B <sub>22</sub>

A<sub>12</sub>는 집단1의 데이터가 집단2로 잘못 분류된 수, 올바르게 분류된 데이터의 수는 (B<sub>11</sub>+B<sub>22</sub>), 잘못 분류된 데이터의 수는 (B<sub>12</sub>+B<sub>21</sub>)이다.

분류 모형의 정확도(accuracy)는 전체 데이터의 수 중 올바르게 분류된 수의 비율로 정의한다. 오류율(error rate)은 전체 데이터의 중에서 오분류된 수의 비율로 정의한다.

$$\text{정확도(accuracy)} = \frac{B_{11} + B_{22}}{n}$$

$$\text{오류율(error rate)} = \frac{B_{12} + B_{21}}{n} \quad (\text{식 7})$$

일반적으로 분류모형은 정확도를 최대화하거나 오류율을 최소화하는 알고리즘을 찾기 위해 노력한다. 여러 분류 모형의 비교 및 평가를 위해서는 리프트차트, ROC 그래프 등을 이용한다.

## 2. 변수선정

자산운용사가 투자예정 펀드 포트폴리오에 개별펀드의 위험등급을 5개 등급으로 구분하여 개별펀드의 투자위험을 투자자들이 직관적으로 이해할 수 있도록 하고 있다. 이러한 위험판단의 기준이 되는 펀드의 위험등급을 고위험

(1)과 저위험(0)으로 더미변수화하여 종속변수로 하였다.

독립변수는 금융투자협회에서 공시하고 있는 국내외 부동산펀드 자산운용사 및 판매사별 부동산펀드 판매 현황의 항목들을 독립변수로 사용하였다.

세부 독립변수 항목으로는 금융투자협회의 국내외 부동산펀드별 실적데이터 중 순자산총액, 보유자산 금액(주식, 채권, 현금성, 기타를 금액과 비중으로 구분), 기준가격, 1개월 수익률, 6개월 수익률, 1년 수익률, 설정일 수익률, 운용보수, 판매보수, 판매수수료를 독립변수로 하여 펀드의 위험을 예측하였다.

〈표 4〉 변수의 정의

변수명	순자산 총액	보유자 산금액 (주식, 금액)	보유자 산금액 (주식, 비중)	보유자 산금액 (채권, 금액)	보유자 산금액 (채권, 비중)	보유자 산금액 (현금성, 금액)
변수명 (영문)	TESSET	STOCK	STOCK@	BOND	BOND@	CASH
변수명	보유자 산금액 (현금, 비중)	보유자 산금액 (기타, 금액)	보유자 산금액 (기타, 비중)	기준 가격	수익률 (1개월)	수익률 (6개월)
변수명 (영문)	CASH@	ESC	ESP@	SP	6MR	1YR
변수명	수익률 (1년)	설정일 수익률	운용 보수	판매 보수	판매 수수료	
변수명 (영문)	1YR	SR	OP	REP	FEES	
변수 설명	* 위험등급에 따른 위험가치 판단: 펀드의 위험등급 1등급(변동성25%이상)을 위험(1), 2-5등급을 저위험(0)으로 구분하여 종속변수로 하였다. * 단위 : 억 원, %					

〈표 5〉 기초통계량 분석

구분	위험	순자산	주식금액	주식비중	채권금액	채권비중	현금금액	현금비중	기타금액	기타비중	기준가격	1개월	6개월	1년	설정일	운용	보수	수수료
평균	.43	165.13	34.83	4.49	0.0000	0.0000	21.68	3.60	165.30	91.90	1067.32	.38	1.27	4.59	19.30	.59	.51	.19
표준편차	.49	622.04	237.30	16.21	0.0000	0.00000	116.99	13.82	817.28	22.40	688.70	3.53	15.55	16.89	34.35	.36	.47	.44
최소값	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-9.00	-5.00	343.26	-5.03	-11.35	-84.08	-83.08	0.00	0.00	0.00
최대값	1.00	5820.00	3000.00	101.00	0.00	0.00	1107.00	100.00	8291.00	100.00	6268.96	59.89	283.99	273.77	136.36	1.10	1.50	2.00

자료 : 금융투자협회(2014.12~2017.11기준 데이터, 402N).

### 3. 분석 결과

#### 1) 기초통계량 분석

데이터마이닝 분류분석을 통한 국내의 부동산펀드의 위험(1)에 영향을 미치는 독립변수들의 주요 기초통계량은 〈표 5〉와 같다. 분석결과는 모형전체의 평균, 표준편차, 최대값, 최소값을 나타낸다. 분류분석을 위한 명목변수인 위험의 평균과 표준편차는 0.43과 0.49이었다. 기타 변수(평균, 표준편차) 분석결과는 순자산(165.13, 622.04), 주식비중(4.49, 16.21), 현금비중(3.60, 13.82), 기타비중(91.90, 22.40), 1개월(0.38, 3.53), 6개월(1.27, 15.55), 1년(4.59, 16.89), 설정일(19.30, 34.35), 운용(0.59, 0.36), 보수(0.51, 0.47), 수수료(0.19, 0.44)였다.

#### 2) 검증모형

##### (1) 인공신경망(ANN) 모형

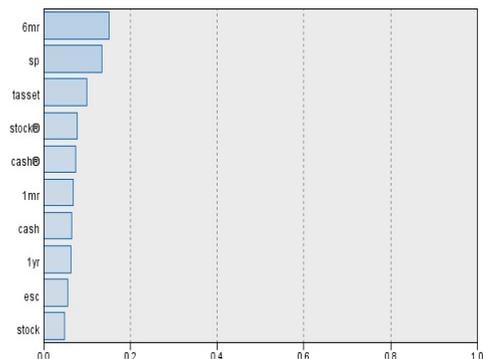
예측분석의 분류의 정확도는 다음 〈표 6〉과 같이 86.82%로 높은 정분류율을 나타냈다. 따라서 인공신경망 모형(ANN)을 활용하여 부동산펀

드의 위험(risk)을 예측할 수 있는 설명력이 충분히 확보되었다고 할 수 있다.

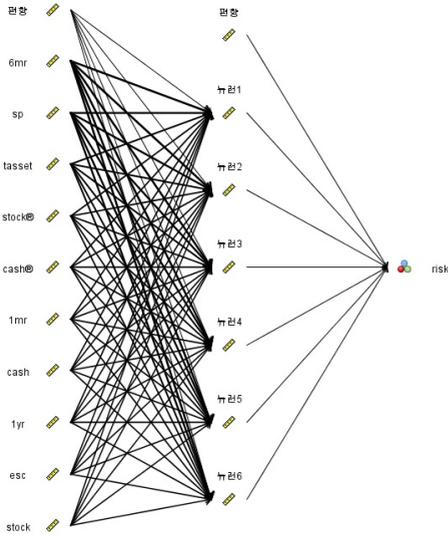
ANN을 활용하여 부동산펀드의 위험(risk)을 예측한 결과 예측변수의 중요도는 6mr(6개월수익률)>sp(기준가격)>tasset(순자산총액) 순으로 중요도가 높은 것으로 분석되었다.

〈표 6〉 정분류율

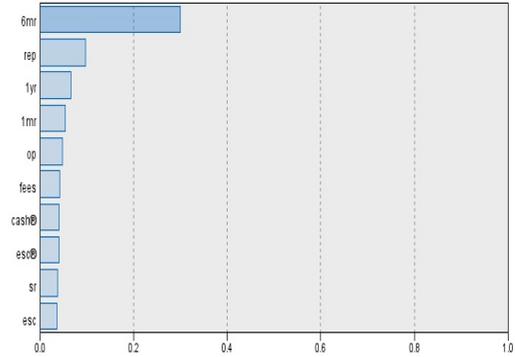
정확	349	86.82%
틀림	53	13.18%
합계	402	



〈그림 3〉 예측변수의 중요도



〈그림 4〉 네트워크



〈그림 5〉 예측변수의 중요도

(2) 서포트벡터머신(SVM) 모형

예측분석의 분류의 정확도는 다음 〈표 7〉과 같으며 70.9%라는 높은 정분류율을 나타냈다. 따라서 SVM 모형을 활용하여 부동산펀드의 위험(RISK)을 예측할 수 있는 설명력이 확보되었다고 할 수 있다.

아울러 SVM을 통하여 예측자의 중요도를 분석한 결과 펀드 위험(RISK)에 영향을 미치는 중요도 분류분석 결과 6mr(6개월수익률), rep(보수), 1yr(1년수익률) 순으로 위험(RISK)이라는 종속변수를 설명하는 중요한 변수임이 분석되었다.

〈표 7〉 정분류율

정확	285	70,9%
틀림	117	29,1%
합계	402	

(3) 로지스틱회귀분석(LR) 모형

0단계의 분류 정확도는 57.2%이며 유의수준 0.004, Wald 8.310, Exp(B) 0.748 이었다. 1단계 분석 결과 전체의 분류의 정확도는 다음 〈표 8〉과 같으며 65.92%의 정분류율을 보였다. exp(b) > 1 일 때 독립변수가 1 단위 증가하면 종속변수의 큰 값이 발생할 가능성이 증가한다. 이러한 예측의 정확성을 바탕으로 종속변수에 미치는 중요도는 stock비중(주식비중), esc비중(기타비중), cash비중(현금비중), 1mr(1개월수익률) 순으로 분석되었다.

〈표 8〉 정분류율

정확	265	65,92%
틀림	137	34,08%
합계	402	

〈표 9〉 예측변수의 중요도

모수 추정값

risk <sup>a</sup>	B	표준오차	Wald	자유도	유의확률	Exp(B)	Exp(B)에 대한 95% 신뢰구간	
							하한	상한
1.0	절편	-163.216	83.351	3.834	1	.050		
	tasset	.001	.002	.528	1	.467	1.001	.998 1.004
	stock	-.003	.002	1.942	1	.163	.997	.993 1.001
	stock <sup>Ⓜ</sup>	1.670	.836	3.991	1	.046	5.312	1.032 27.344
	bond	0 <sup>b</sup>	.	.	0	.	.	.
	bond <sup>Ⓜ</sup>	0 <sup>b</sup>	.	.	0	.	.	.
	cash	.006	.004	2.844	1	.092	1.006	.999 1.014
	cash <sup>Ⓜ</sup>	1.642	.832	3.892	1	.049	5.165	1.011 26.396
	esc	-.001	.001	.440	1	.507	.999	.998 1.001
	esc <sup>Ⓜ</sup>	1.657	.834	3.950	1	.047	5.242	1.023 26.853
	sp	-.003	.001	16.040	1	.000	.997	.995 .998
	1mr	.167	.053	9.872	1	.002	1.182	1.065 1.312
	6mr	.012	.020	.324	1	.569	1.012	.972 1.053
	1yr	.017	.019	.779	1	.377	1.017	.980 1.056
	sr	.004	.004	1.212	1	.271	1.004	.997 1.011
	op	.101	.466	.047	1	.828	1.106	.444 2.757
	rep	.155	.309	.251	1	.616	1.168	.637 2.140
	fees	-.276	.289	.912	1	.339	.759	.431 1.337

a. 참조 범주는 1.0입니다.

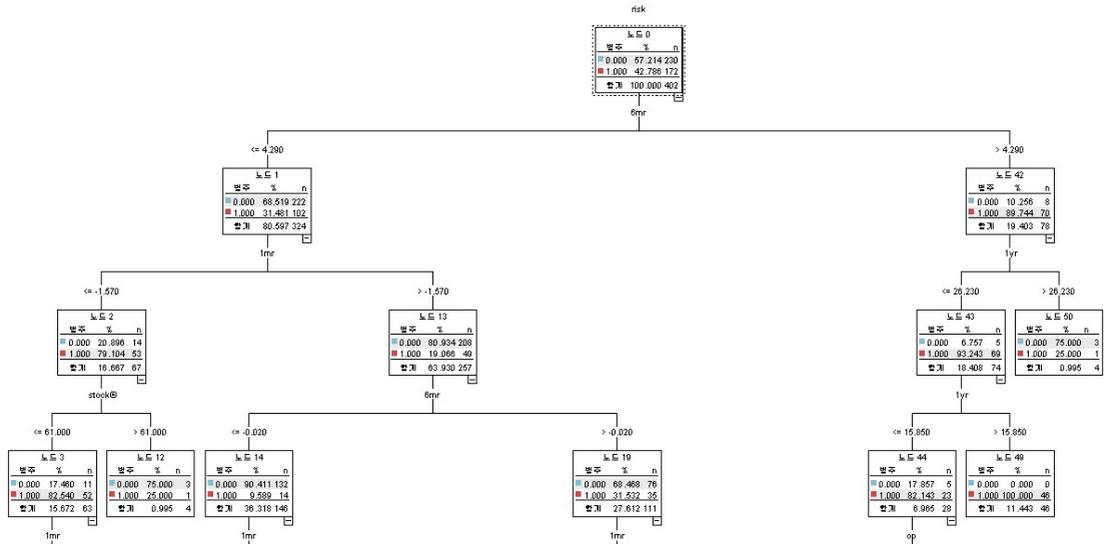
b. 이 모수는 중복되었으므로 0으로 설정됩니다.

(4) 의사결정나무(C5.0) 모형

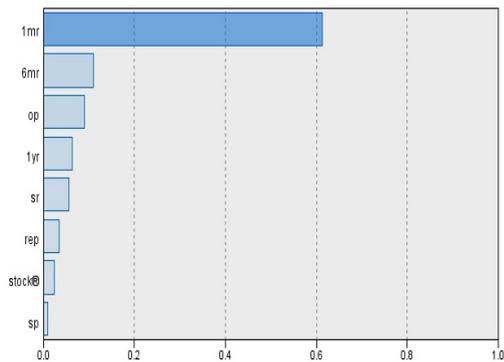
전체의 분류의 정확도는 다음 〈표 10〉과 같으며 97.76%의 높은 정분류율을 나타냈다. 이러한 예측의 정확성을 바탕으로 종속변수에 미치는 중요도는 1mr(1개월수익률), 6mr(6개월수익률), op(운용보수) 순으로 분석되었다.

〈표 10〉 정분류율

정확	393	97.76%
틀림	9	2.24%
합계	402	



〈그림 6〉 의사결정나무 노드



〈그림 7〉 예측변수의 중요도

〈표 11〉 정분류율과 변수의 중요도 비교

구분	ANN	SVM	LR	C5.0
정분류율	86,82%	70,9%	65,92%	97,76%
예측변수의 중요도	6mr) sp) tasset	6mr) rep) 1yr	esc) 1mr	stock) esc) cash

위의 결과를 종합해 보면 부동산펀드의 위험 예측에 가장 큰 영향을 미치는 요인은 전체의 분류의 정확도는 의사결정나무 모형(C5.0)이 펀드 위험 예측을 위한 가장 높은 정분류율(97.76%)을 나타냈다. 이를 바탕으로 국내의 부동산펀드 위험 예측을 위한 가장 큰 영향을 미치는 독립변수는 1개월수익률(1MR), 6개월수익률(6MR), 기준가격(OP) 순으로 위험을 설명하는 변수의 중요도가 높은 것으로 나타났다.

### 3) 분석결과

분류분석을 통한 연구모형간 분류분석을 통한 위험예측 결과는 다음 〈표 11〉과 같다.

## V. 결론

데이터마이닝을 이용한 국내외 부동산펀드 판매결과를 분류분석을 통해 위험 예측분석한 결과 의사결정나무모형(C5.0)이 부동산펀드 위험 예측에 가장 높은 정분류율(97.76%)로 분류의 정확도를 보였다. 이러한 높은 정분류율을 바탕으로 종속변수인 부동산펀드 투자의 고위험(1)을 예측하는 중요한 독립변수는 1개월수익률(1MR), 6개월수익률(6MR), 기준가격(OP), 1yr(1년수익률)순으로 위험을 설명하는 변수의 중요도가 높은 것으로 나타났다.

실제 판매된 국내외 부동산관련 펀드상품 판매 결과를 데이터마트로 구성하여 실증분석에 활용하였다는 점에 시사성이 크다 하겠다. 부동산금융에도 기계학습 모델링을 통한 알고리즘 분석을 통해 빅데이터 예측분석을 결과물로 도출하였다는 점이다. 이를 통해 위험의 특성을 분석하였고 기계학습 모델링을 통하여 예측분석을 실증 하였다. 또한 연구모형을 실증에 거시지표 등 외부 요인이 아니라 실제 투자가 이루어진 투자모형내의 변수들을 통하여 확인하였다는 점이 연구의 큰 성과라 할 수 있다.

자산운용사 등의 부동산펀드 상품의 판매결과를 바탕으로 분류분석 결과를 통한 예측력을 바탕으로 위험 또는 불확실성을 분류하고 예측할 수 있다면 위험대비 수익성을 극대화할 수 있으리라 본다.

부동산펀드는 환금성이 높은 반면 투자수익이 안정적이지 못하다는 한계를 가지고 있다. 본 연구의 데이터마이닝을 이용한 기계학습을 통한

예측력 확보와 투자성과 판단에 영향을 미치는 중요변수 도출이 이러한 한계를 극복하는데 자산운용사 및 판매사 부동산펀드 투자 운용담당자들에게 마케팅 분석시 기준을 제시하였다.

자산운용사 및 펀드판매사의 부동산펀드 위험 가치 판단에 펀드판매 세부 항목들 중 운용보수가 펀드 투자 등 위험 판단에 가장 중요한 요인이라는 점을 실증하였다는 점에서 시사하는 바가 크다 할 것이다.

본 연구는 데이터마이닝을 통한 예측분석과 중요변수 선정에 있어 독립변수를 다양하게 선정하여 예측의 설명력을 극대화 시키지 못하였다. 국내 부동산펀드 투자 위험과의 단순 비교에 그치고 세밀한 계량화된 비교실증을 통해 부동산펀드의 특성을 도출하지 못했다는 점은 연구의 한계라 할 것이다. 향후 연구에서는 이점을 보완하여 국내외 부동산투자의 특성을 데이터마이닝을 통하여 예측하고 비교 분석하고자 한다.

## 참고문헌

1. 김경민 외 2인, 2015, 「국민연금 기금운용을 통한 부동산투자 특성에 관한 연구」, 『부동산학보』, 제 62권.
2. 김나영 · 이은경, 2014, 「나무구조의 분류분석에서 변수 중요도에 관한 고찰」, 『응용통계연구』, 27권 5호.
3. 김병준 · 이창석, 2016, 「미국 REITs시장과 주식 시장 수익률간의 상호 영향력 검증」, 『부동산학보』, 제66권.
4. 김영곤 외 4인, 2014, 『부동산금융과 투자(제14판)』, 부연사.

5. 김홍년, 2015, 「글로벌 리츠시장 현황 및 리츠산업의 활성화 필요성」, 『리츠저널』, VOL.16.
6. 남연우 · 고석찬, 2015, 「부동산펀드 성과 영향요인 분석」, 『부동산학보』, 제62권.
7. 박덕배, 2015, 「부동산 간접투자시장의 특징과 시사점」, 『VIP REPORT(15-7호)』, 현대경제연구원.
8. 박창이 외 4인, 2015, 『R을 이용한 데이터마이닝』, 교우사.
9. 방경식 · 장희순, 2016, 『부동산학총론』(제4판), 부연사.
10. 여수진 외 2인, 2014, 「리츠 수익률의 결정요인에 관한 연구」, 『대한부동산학회지』, 제32권 제1호.
11. 이병훈 외 3인, 2011, 「회귀분석을 통한 부동산투자회사(REITs) 오피스빌딩의 수익률 영향요인 분석」, 『대한건축학회논문집』, 제27권 제9호.
12. 이안위튼 · 아이베프랑크 · 마크홀, 2013, 『데이터마이닝(기계학습의 이론과 응용)』, 에이콘.
13. 이재길, 2016, 『R프로그램에 기반한 다변량분석 및 데이터마이닝』, 황소걸음아카데미.
14. 이치주 외 1인, 2010, 「리츠와 건설경기, 부동산경기, 주식시장과의 관계분석」, 『한국건설관리학회지』, 제11권 제5호.
15. 전해정, 2013, 「자산가격결정이론에 기반한 주택가격결정요인 분석에 관한 연구」, 『부동산학보』, 제52권.
16. 조재영, 2016, 「프로젝트 금융의 안정성 확보에 관한 연구」, 『부동산학보』, 제64권.
17. 조현민 외 1인, 2011, 「한국 부동산시장 및 자본시장과 부동산투자회사(REITs) 간의 연관성 분석」, 『국토연구』, 71권.
18. 최치순, 2014, 「REITs 수익률의 ARIMA Model 설정에 관한 연구」, 『부동산학보』, 제58권.
19. 최효비 외 1인, 2016, 「은퇴계층의 부동산자산 운용에 관한 결정요인 분석」, 『부동산학보』, 제65권.
20. 캐빈머피, 2015, 『머신러닝(Machine Learning)』, 에이콘.
21. 홍아름 외 2인, 2010, 「데이터마이닝 기법을 이용한 서울시 오피스빌딩 투자특성 예측에 관한 연구」, 『서울도시연구』, 제11권 제2호.
22. Christopher Grandrud, 2015, *Reproducible Research with R and RStudio*(second edition), CRC Press.
23. David. Enke · Suraphan Thawomwong, 2005, “The use of data Mining and neural networks for forecasting stock market returns” , *Expert Systems with Application*, Vol.29.
24. Fn-Guide(<http://www.fnguide.com/>)Index During The 2000-2011 Time Period” , *Journal of Business & Economics Research*, 12(1), 2014.
25. K.C.Tseng, 2010, “Forecasting Financial Stocks Using Data Mining” , California State University Fresno, Forecasting.
26. Tilo Wendler · Sören Gröttrup, 2016, *Data Mining with SPSS Modeler*.
27. 금융투자협회(<http://dis.kofia.or.kr/>)

논문접수일 : 2017년 10월 16일  
 심사(수정)일 : 1차 2017년 11월 13일  
                   2차 2017년 11월 28일  
 게재확정일 : 2017년 11월 30일

## 국문초록

본 연구는 국내외 부동산펀드 상품군의 위험 특성을 분석하였다. 위험특성 분석으로 데이터마이닝을 통한 분류분석을 실시하였다. 분류분석을 통한 위험특성 예측을 통해 자산운용사 및 판매사의 상품기획 및 자산운용 담당자들에게 국내외 부동산펀드 투자상품 및 투자항목 선택 기준을 제시하였다. 이를 통해 위험을 최소화하고 수익성을 제고 시키고자 하는데 주된 연구의 목적을 두었다. 국내외 부동산펀드 판매결과에 대한 분류분석은 402개의 자산관리회사 성과 데이터를 원데이터(2014.12~2017.11)로 활용한 데이터마트로 분석을 실시하였다. 분류분석 결과 위험에 영향을 미치는 변수의 중요도를 확인하였다. 분류분석 결과 정분류율은 의사결정나무모형(C5.0)의 97.76%로 가장 높게 분석되었다. 부동산펀드 투자의 고위험(1)을 예측하는 중요한 독립변수는 1개월수익률(1MR), 6개월수익률(6MR), 기준가격(OP), 1yr(1년수익률) 순으로 위험을 설명하는 변수의 중요도가 높은 것으로 나타났다.

주제어 : 국내외 부동산펀드, 위험등급, 위험예측, 분류분석, 정분류율

