

# 데이터마이닝을 이용한 글로벌 리츠 투자 특성 예측

Forecasting of Investment Characteristics of Global REITs Using Data Mining

김 경 민\* · 김 천 규\*\*

Kim, Kyeong Min · Kim, Cheon Kyu

## 目 次

- |                      |                  |
|----------------------|------------------|
| I. 서 론               | 2. 변수선정 및 기초통계분석 |
| 1. 연구의 배경 및 목적       | 3. 분석 결과         |
| 2. 연구의 범위 및 방법       | IV. 결 론          |
| II. 선행연구 검토 및 이론적 기초 | 1. 연구의 결과 요약     |
| 1. 선행연구 검토           | 2. 연구의 시사점 및 한계  |
| 2. 대체투자의 이론적 기초      | <abstract>       |
| III. 실증분석 및 결과       | <참고문헌>           |
| 1. 검증모형 및 분류모형 평가    |                  |

## ABSTRACT

### 1. CONTENTS

#### (1) RESEARCH OBJECTIVES

The results of the analysis of the prospectus and investment characteristics of the global barrel re-indebted fuels will be used by institutional investors and investment company practitioners who want to invest in overseas REITs to select overseas investment region selection and investment items And to contribute to the minimization of risk by providing assistance to the public.

#### (2) RESEARCH METHOD

The data used for valuation analysis and classification analysis of REITs Investment consisted of 332 asset management companies' performance data on real estate overseas investment-type REITs investment products, The investment profit ratio of overseas investment-type commodities was 2016. It was 1 year immediately before the end of October 2015 (from May 1, 2015 to October 20, 2010).

#### (3) RESEARCH FINDINGS

We use data mining to classify and analyze the investment characteristics of the Global Litigation Re-Induced Firm by securing predictive power using these data mining methods.

\* 주저자 : 강남대학교 부동산학과 시간강사, ㈜폴드엔코리아리얼티 대표, 부동산학 박사, kkmin0390@naver.com

\*\* 공동저자 : 강남대학교 부동산학과 박사과정, kimchg87@hanmail.net

▷ 접수일(2016년 11월 29일), 수정일(1차 : 2017년 2월 9일), 게재확정일(2017년 2월 10일)

## 2. RESULTS

Results Characteristics of Investment in Global Ritz Re-Indirect Fund Using Data Mining As a result of machine learning for predictive analysis, the neural network model has the highest positive classification rate (99.7% accuracy) for predicting investment outcome Indicated. Next, we showed the prediction accuracy (Chon classification rate 98.82%) for the significance analysis of prediction variables of decision tree model (C 5.0) and logistic regression model. Based on the accuracy of such prediction, the independent variables affecting the investment result (IV) of the dependent variable as a result of the importance analysis of the variable are OM (operational maintenance), total net asset (NAV), price standard (BP) I was able to confirm the fact that.

## 3. KEY WORDS

- Real Estate Indirect Investment, Global REITs, Investment Value, Forecast Analysis, Sorting Rate, Importance

## 국문초록

글로벌 금융위기 및 유럽 재정위기, 중국의 경기둔화 등 대외적 요인과 국내경기 둔화, 전세시장 역전 등의 대내적 요인에 따른 부동산 경기 부진으로 거래가 매우 침체되어 있다. 리츠(Real Estate Investment Trusts, REITs)와 부동산펀드(Real Estate Fund, REF) 등을 통해 임대수익이나 매각수익을 향유하는 부동산 간접투자시장이 급성장하고 있다. 리츠투자는 글로벌 포트폴리오 및 대체투자 다변화, 수익률 극대화라는 가치하에 연기금 등 기관투자자의 수임을 받은 자산운용사를 중심으로 판매사를 통해 지속적으로 재간접펀드 일종인 글로벌리츠재간접펀드에 글로벌부동산펀드와 더불어 투자자산을 증대시키고 있다. 글로벌리츠재간접투자의 예측분석 및 투자특성 분석결과를 바탕으로 투자하고자 하는 기관투자자 및 투자회사 실무담당자에게 투자 특성 예측분석을 통해 해외 투자지역 선별 및 투자항목 선택에 조력을 줌으로써 위험을 최소화하는데 기여하고자 하는 목적에서 연구를 실시하였다. 데이터마이닝을 통한 분류분석을 위한 데이터마이닝 투자가치 예측분석 결과 예측분석의 분류의 정확도는 99.7%로 높은 정분류율을 나타냈다. 따라서 인공지능경망 모형(ANN)을 활용한 분석결과를 바탕으로 자산운용사 및 판매사의 상품기획 및 자산운용 담당자들에게 글로벌 리츠상품 개발 기준 및 판매 기준 등을 제시해 줄 것을 기대한다.

**핵심어 : 부동산간접투자, 글로벌리츠, 투자가치, 예측분석, 정분류율, 중요도**

## I. 서론

### 1. 연구의 배경 및 목적

글로벌 금융위기, 유럽 재정위기 등 대내외적 경기 침체 요건 하에 부동산 거래 또한 매우 침체되어 있다.

해외 부동산에 투자하는 글로벌 리츠(REITs · 부동산투자신탁) 펀드에 관심이 쏠리고 있다. 국내외 증시 흐름이 불안한 가운데 부동산 간접투

자 상품인 리츠펀드는 주식형펀드보다 안정적인 수익을 기대할 수 있다.<sup>1)</sup>

주식과 채권 모두 투자전망이 불확실한 상황에서 대체투자수단으로 부동산간접투자가 연기금 등 기관투자자들을 중심으로 투자가 집중되고 있다.<sup>2)</sup> 부동산간접투자 상품인 글로벌리츠 또한 글로벌 대체투자 다변화 및 수익성 제고에 따라 투자자산배분 또한 꾸준히 증대되고 있다.

부동산간접투자의 증대와 자본 시장에서의 강력한 영향력으로 인한 공개시장의 공개요구로 언론

1) 한국부동산신문, "글로벌리츠펀드에 투자해볼까", 2016.06.17자 기사(<http://부동산신문.com>)

2) 박덕배, "부동산 간접투자시장의 특성과 시사점", *VIP REPORT*, 현대경제연구원, 2015, 15-7호, p.7.

의 불균형 또한 심화되고 있다.<sup>3)</sup> 그럼에도 부동산 간접투자인 부동산펀드와 리츠는 연기금 등 기관투자자의 위험 대비 수익률이 높은 관계로 해외상품에 대한 투자할당을 늘리고 있는 추세이다.

연초 미국 기준 금리 인상에 대한 우려가 꾸준히 제기되고 있기는 하나 인상이 지연되면서 리츠 수익률이 증대되고 있다. 각국의 저금리 정책 등에 힘입어서도 글로벌 부동산 시장이 살아나고 있으며 글로벌리츠펀드 수익률도 호전되고 있다.

국내 자산운용사의 글로벌리츠펀드 투자는 연기금 등 기관투자자들의 위탁운용 방식을 통하여 재간접투자라는 방식을 주로 사용하고 있다.

자산운용사 등을 통해 글로벌리츠 투자가 해외 여러 투자처로 증대되고 있음에도 불구하고 기존의 리츠 관련 연구는 주로 국내 리츠와 실물부동산 그리고 거시경제변수간의 관계를 분석하거나 리츠 투자수익률 결정요인 분석에 집중되었다.

이에 본 연구는 글로벌리츠 재간접 투자를 포커스로 하여 예측분석 및 투자특성 분석을 실시하였다. 이를 통해 글로벌리츠펀드에 투자하고자 하는 기관투자자 및 투자회사 실무담당자에게 예측분석과 투자 항목 중요도 분석을 바탕으로 글로벌리츠 투자지역 선별에 조력을 줌으로써 수익성을 제고하고 리스크를 최소화 시키는데 기여함을 주된 연구 목적으로 하였다.

## 2. 연구의 범위 및 방법

최근 1년간의 글로벌리츠재간접 투자사례(원데이터)를 바탕으로 투자수익성의 투자가치를 투자성과 여부로 구분하여 예측을 실시코자 하였다.<sup>4)</sup> 이러한 투자가치 판단을 위해 정제된 데이터마트를 구성하여 데이터마이닝을 활용한 기계학습 예측분석을 실시한다. 이러한 데이터마

이닝을 활용해 예측력을 확보하여 글로벌리츠재간접의 투자특성을 분석한다.

국내 자산운용사가 판매사를 통하여 판매된 재간접투자상품의 해외 리츠투자를 통한 투자성과 예측을 위하여 데이터마이닝을 이용한 기계학습 분석으로 의사결정나무, 인공신경망, 서포트 벡터머신, 로지스틱회귀분석을 실시하였다.

선행연구 및 기초이론검토를 통하여 리츠 투자 관련 선행연구를 분석하여 본 연구와의 차별성을 부각시킨다. 또한 자산운용사의 글로벌리츠 재간접 투자 현황 및 투자수익률을 확인하여 대체투자로서 부동산투자회사(REITs)의 투자성과를 평가한다.

실증분석은 국내 자산운용사 및 판매사의 글로벌리츠펀드재간접 투자 실적을 원데이터로 하여 데이터마이닝을 이용한 투자 특특성 예측 분석을 실시한다.<sup>5)</sup> 이를 위해 분류의 정확도 분석을 실시하고 정분류율을 통한 예측력 확보를 바탕으로 종속변수인 투자성과에 중요한 영향을 미치는 독립변수를 검증한다.

리츠 투자가치 예측분석 등을 위하여 사용된 로데이터는 글로벌리츠재간접펀드 투자상품 332개의 자산운용사 및 판매사들의 실적데이터를 로데이터로 구성하였다. 글로벌리츠펀드 상품들의 예측모형 판단의 기간은 최근 1년간의 기준을 2016.10월말로 판단하고 직전 1년(2015.11.01~2016.10.30자)으로 하였다.

분석 데이터 마트 구성을 위한 로데이터는 금융투자협회 데이터를 참조하였다. 글로벌리츠재간접펀드와 국내 리츠와의 비교를 위하여 FnGuide Index로 부터 국내 리츠수익률 자료(케이탑, 광희, 트러스 7호)를 활용하였다.

데이터마이닝을 위하여 SPSS22.0, SPSS

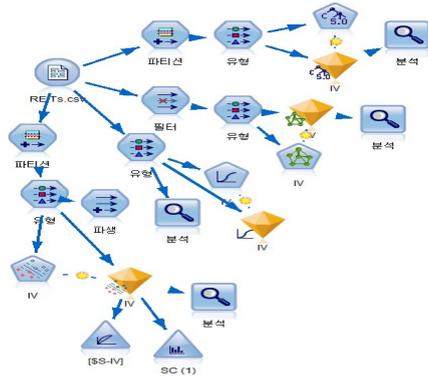
3) Scott R. Muldavin, *The New World of Real Estate Finance*, Real Estate Finance, The Muldavin Company, Inc., 2001.8, p.75.

4) 글로벌리츠재간접펀드란 재간접 투자기구로서 각종 증권거래소에 상장된 리츠(REITs)에 투자하여 일정수준의 지속적인 임대수익과 리츠 가격상승에 따른 자본차익을 추구하며 또한 부동산 관련 투자를 통한 일정부분 인플레이션 헷지를 추구하는 상품을 말한다.

5) 데이터마이닝을 명확히 정의하기는 어려우나 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, R을 이용한 데이터마이닝, 교우사, 2015, p.3.)

Modeler, R-program을 활용하였다. 실증분석 연구모형은 다음 <그림1>과 같다.

<그림1> 데이터마이닝 분석 모형



## II. 선행연구 검토 및 이론적 기초

### 1. 선행연구 검토

기존 선행연구는 리츠수익률 결정요인분석과 리츠와 다른 거시경제지표 등과의 관계에 관한 연구가 주류를 이루었다.

국내의 리츠에 대한 투자 수익률 관련한 선행연구는 다음과 같다.

국내 선행연구로는 허필원 외 3명(2012)<sup>6)</sup>은 해외리츠펀드의 투자지역별 분류와 외부거시경제변수와의 상호연관성을 분석하였다. 해외리츠펀드 3개의 상호연관성을 분석하기 위하여

VAR모형 기반으로 한 검정을 실시하였다.

유상철(2012)<sup>7)</sup>은 2001년부터 2010년 말까지 설립된 62개 리츠 중 공모로 모집되어 상장된 13개 리츠를 대상으로 하였다. 부동산의 용도, 위치, 자산의 규모, 부동산의 규모, 자산관리 수수료 비율이 성과에 영향을 미치는 것으로 나타났다.

이경순(2010)<sup>8)</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. 원필재(2010)<sup>9)</sup>는 2010년 운영 중인 25개 부동산 투자회사를 대상으로 하였으며, 주당순이익, 재무현금흐름, 현금흐름, 매출액, 영업이익률, 고정자산회전율이 배당률에 영향을 미친다고 하였다.

김재훈(2008)<sup>10)</sup>은 리츠의 자기자본수익률의 지표인 배당률과 유의한 관계가 있는 독립변수로 LTV, 연간이자율, 자산관리회사, 부동산 종류, 리츠의 종류임을 실증하였다.

장용삼외 2인(2011)<sup>11)</sup>은 리츠와 타 금융자산의 수익-위험 분석을 통해 리츠의 특성을 확인하였다.

해외 선행연구에서는 K.C.Tseng(2010)<sup>12)</sup>이 '데이터마이닝을 활용한 금융주식 예측'연구에서 주가의 예측을 위한 예측의 정확성을 증가시키기 위해 인공지능(NN)과 OLS모형을 비교하였다.

David Enke(2005)<sup>13)</sup>는 데이터 마이닝을

6) 허필원 · 허정석 · 김경수 · 정미나, "해외리츠펀드와 거시경제변수의 상호연관성에 관한 연구", 부동산금융논문 현상공모전 수상작 모음집, 메리츠증권, 부동산금융연구소, 2011, pp.159~198.

7) 유상철, "리츠의 운영성과와 특성", 강원대학교 박사학위논문, 2012, pp.1~131.

8) 이경순, "임대형부동산 간접투자상품의 배당수익률 결정요인분석", 건국대학교 석사학위논문, 2010, pp.1~185.

9) 원필재, "재무분석을 통한 부동산투자회사 배당률 영향요인 분석", 단국대학교 석사학위논문, 2010, pp.1~93.

10) 김재훈, "리츠의 자기자본수익률에 영향을 미치는 요인에 관한 실증 연구", 연세대학교 석사학위논문, 2008, pp.1~66.

11) 장용삼 · 김진호 · 김해중 ·, "부동산투자신탁의 상관성에 관한 연구", 한국지적학회지, 한국지적학회, 2008, pp.95~108.

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

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

위한 기계학습에 사용되는 정보획득 방법을 소개하였다. 다수의 금융 및 경제변수의 예측 관계를 평가하였다. 수준 추정 및 분류를 위한 신경망 모델이 효과적인 미래 가치 예측을 위해 조사되었다.

본 연구의 기존 선행연구와의 차별성은 자산 운용사 및 판매사를 통하여 실무에서 투자된 부동산 해외 리츠투자 상품의 실적 데이터를 바탕으로 투자성과를 예측하고 투자를 결정하는 요인을 데이터마이닝을 이용하여 분석하였다는 점이다.

## 2. 이론적 기초

### 1) 리츠의 정의

리츠는 다수의 투자자로부터 자금을 모집하여 부동산 및 부동산과 관련된 유가증권에 투자·운영하여 발생한 수익을 투자자들에게 배당하는 부동산 간접투자방식이다. 내용적 측면에서 부동산간접투자 기구라 할 수 있다. 법적 측면에서는 부동산 간접투자의 기구라고 할 수 있다.

### 2) 리츠의 유형 및 가치평가

#### (1) 리츠의 유형

리츠의 주요한 3가지 유형으로는 지분형(Equity Trusts), 저당대출형(Mortgage Trusts), 혼합형(Hybrid Trusts)이 있다. 초창기에는 지분형이 일반적이었다. 1970년대 중반에는 저당대출형이 더 중요하게 되었다. 최근 들어 지분형이 다시 더욱 중요해지고 있으며, 회사의 수나 자본규모에 있어서 지배적인 유형이 되었다. 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 발전

하였는데, 이를 혼합형 펀드라고 한다.<sup>14)</sup>

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

### (2) 리츠의 가치평가 방법

펀드수익률은 운영성과를 나타내는 지표로 일정기간 동안 기준가격이 얼마나 증감했는지를 나타낸다.<sup>15)</sup> 리츠가 공시하는 자료들은 다양한 정보를 제공한다. 이러한 자료들에는 리츠들이 소유하고 있는 부동산가치를 정확하게 제공해 주지 않는다. 이와 관련된 합리적인 해결책은 일반인이 생각하는 것보다 어려운 작업이다.<sup>16)</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)와 배당성장율(G)의 차로 나눈 값이다.

## 3) 글로벌리츠 투자의 특성

### (1) 글로벌리츠의 현황

미국에서는 부동산 관련 간접투자기구로서 리츠와 부동산뮤추얼펀드(Real Estate Mutual Fund)가 있다.

리츠는 1993년 증권법에 의해 규유되었다. 투자상품으로서 리츠의 성격은 '중위험·중수익'의 상품이다.

수익성 측면에서는 주식보다는 적은 반면 채권보다는 수익성이 높다. 위험의 측면에서도 주

14) 김영권의 4인, 부동산금융과 투자, 부연사(제14판), 2014, pp.745~746.

15) 방경식·장희순, 부동산학총론(제4판), 부연사, 2016.08, p.580.

16) 김영권의 4인, 상계서, 2014, p.114.

식보다는 적고 채권보다는 높은 투자처로 인식되고 있다. 그렇기 때문에 기관투자자들이 자산 포트폴리오 차원에서 위험 헤지 투자 상품으로 이용하고 있다.<sup>17)</sup>

대체투자로서 부동산에 투자하는 글로벌리츠재간접은 2016년 6월 기준 최근 1년 평균 수익률이 4.8%다. 이는 같은 기간 국내 주식형 펀드 수익률이 -4.41%, 해외 주식형 펀드 수익률이 -19.74%로 부진했던 것과 비교된다.<sup>18)</sup>

〈그림 2〉 글로벌리츠재간접펀드 수익률



국내 리츠도 2013년 10월~2016년 10월 까지 49개월의 국내 상장 리츠 중 3개사(광희,케이탑,트러스제7호) 평균 투자수익률과 미국 리츠의 최근3년의 수익률 비교 결과 국내 상장 3개사 평균수익률은 0.77%, 미국다우존스 평균 리츠수익률이 0.47%로 국내상장 3개사의 리츠수익률이 더 높다.<sup>19)</sup> 이를 통해 국내의 자산운용사의 글로벌리츠재간접투자 뿐 아니라 국내 상장 리츠도 꾸준한 성장세를 지속하고 있다는 것을 보여준다. 이를 통해 리츠에 대한 직·간접, 재간접 투자의 열기는 지속될 것으로 본다.

## (2) 글로벌리츠의 특성

글로벌리츠재간접은 펀드 하나로 전 세계 부동산에 간접투자 한다. 주식 및 채권과 다른 성격으로 인해 포트폴리오 분산투자효과가 있다. 상장부동산펀드 및 리츠는 언제든지 환매가 가능한 구조를 가지고 있다.

리츠 펀드는 경기 회복시 임대수요 증가에 따른 수익성 향상 및 보유자산 가치증대가 기대된다. 물가 상승시 임대료 인상을 통해 인플레이션 헤지 효과가 있다. 주식, 채권과 같은 전통자산대비 장기성과는 우수하며, 낮은 상관관계로 분산투자에 적합한 상품이다.

## Ⅲ. 실증분석 및 결과

데이터마이닝은 모형의 타당성 평가를 위하여 데이터를 훈련데이터(training data)와 평가데이터(test data)로 구분한다.<sup>20)</sup> 훈련데이터는 모형을 생성하는데 이용되고, 평가데이터는 최종적인 평가를 수행하는데 사용된다.

평가데이터는 모형의 생성에 전혀 사용되지 않으며 일반화의 검토를 위하여 남겨두는 데이터이다. 본 연구에서는 훈련 데이터 80%, 검증용 데이터 10%, 평가데이터 10%로 나누어 분석을 실시하였다.

본 연구에서는 해외 리츠의 수익률 예측분석을 위하여 분류모델(Classification Model)링통하여 예측의 정확도를 확인하였다. 이러한 예측의 정확성을 바탕으로 판매사별로 해외 리츠 판매된 최근1년 수익률을 사례로(332N)하여 투자성과 특성을 분석한다. 분류모델에 사용된 이론을 살펴본다.

17) 여수진의 2인, "리츠 수익률의 결정요인에 관한 연구", 대한부동산학회지, 대한부동산학회, 2014, 제32권1호, pp.187~188.(제인용)

18) [http://부동산신문.com/print\\_paper.php](http://부동산신문.com/print_paper.php)(펀드평가사 KG제로인)

19) Data Guide program을 통한 국내의 리츠 관련 데이터를 참조하여 저자가 재구성함

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

### 1. 검증모형 및 분류모형 평가

#### 1) 검증모형

##### (1) 로지스틱회귀분석

(Logistic Regression, LR)

로지스틱회귀분석(Logistic Regression, LR)이 선형회귀분석과의 가장 큰 차이점은 타겟 vfd 값의 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>21)</sup>

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

로지스틱 구조는 다음과 같다.

$$F(t) = \frac{\exp(t)}{1+\expa(t)}$$

##### (2) 의사결정나무 분석

(Decision Tree, C5.0)

의사결나무모형은 비선형 회귀분석, 비선형 판별분석의 하나로 분류되는데, Tree모형 그자체로서 분류, 또는 예측과정으로 수행되기도 한다.<sup>22)</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 다지분리알고리즘이다. 엔트로피, 정보이득 개념을 사용하여 분리기준을 결정한다.

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

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

정보이득은 엔트로피의 감소량이며, 엔트로피가 가장 낮아지는 속성을 찾는 것이 목적이다.

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

$Gain(A)$  자료는 자료(S)를 독립변수(A)의 값에 따라  $n$ 개로 나누었을 때 얻게 되는 정보이득이다.<sup>23)</sup>

##### (3) 인공신경망 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 예측하는 비선형구조를 가지는 모형이다.<sup>24)</sup> ANN은 SVM같이 블랙박스 알고리즘이다. 그래서 뉴런의 이동과 데이터 변이의 구조와 메카니즘은 매우 복잡하고 독창적이지 않다.

뉴런의 기능과 활동 구조식은 다음과 같다.

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

신경망에 있어서의 신경은 정보가 통과하는 다층 구조로 하나의 Input layer, 하나 또는 그 이상의 Hidden layer, 그리고 Output layer로

21) Tilo Wendler · Soren Grottrup, *Data Mining with SPSS Modeler*, Springer, 2016, pp.733~736.

22) 이재길, R프로그램에 기반한 다변량분석 및 데이터마이닝, 황소걸음아카데미, 2016, p.321.

23) Begas, BIG DATA, 빅데이터분석전문가 교재, 한국데이터베이스진흥원, 2016, pp.300~307.

24) Begas, 상계서, 2016, p.350.

되어 있다. Input과 Output layer 사이에 하나 혹은 그 이상의 Hidden layer가 존재할 수 있다. 이러한 다계층 신경은 위에서 언급한 바와 같이 전의 계층의 신경으로부터 데이터를 얻는다. 조작된 데이터는 다음의 Hidden layer 혹은 Output layer의 신경에 전달되어졌다.<sup>25)</sup>

(4) 서포트벡터머신 분석(Support Vector Machine, SVM)

SVM(Support Vector Machine)은 비선형분류알고리즘 중 하나로 가장 효율적이고 유연한 방법 중 하나이다. 초월평면에 가장 근접한 인스턴스들을 Support Vector라 부른다. 선형 모델의 최대 단점은 클래스 간의 선형경계 밖에 표현하지 못한다는 점이다.<sup>26)</sup>

예측이 정확하고 여러 가지 형태의 자료에 대하여 적용이 쉽지만 확률 추정 값을 요구하는 분야에는 잘 사용되지 않는다.<sup>27)</sup>

SVM은 서포트벡터 분류를 의미한다. 또한 분류문제뿐 아니라 회귀문제도 적용가능하다.

두 개의 클래스를 분리하는 초월 평면을 다음과 같은 방정식으로 표현할 수 있다.

$$x = w_0 + w_1a_1 + w_2a_2$$

최대 마진 초월평면은 클래스를 가장 많이 분리하는 평면에 해당하며 다음과 같다.

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

2) 분류모형 평가

모형의 평가는 서로 다른 모형 중 어떤 모형이 가장 우수한 예측력을 보유하고는 비교하여 분석하는 과정이다.

두 개의 집단이 있고 n개의 소속집단을 이는

데이터가 있다면 분류모형을 이용하여 각 데이터를 분류한다. 실제집단과 모형에 의해 분류한 집단을 비교하여 그 결과를 아래와 같이 <표 3>으로 요약할 수 있다.

<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>

여기서 B<sub>ij</sub>는 집단 A<sub>i</sub>의 데이터가 집단 A<sub>j</sub>로 분류된 수를 의미한다. A<sub>12</sub>는 집단1의 데이터가 집단 2로 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 수중 올바르게 분류된 수의 비율로 정의한다. 오류율(error rate)은 전체 데이터의 중에서 오분류된 수의 비율로 정의한다.

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

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

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

2. 변수선정 및 기초통계분석

1) 변수선정

리츠수익률은 무위험금융상품인 국고채수

25) Tilo Wendler · Soren Grottrup, Data Mining with SPSS Modeler, Springer, 2016, pp.843~846.

26) 이안위튼 · 아이베프랑크 · 마크홀, 데이터마이닝(기계학습의 이론과 응용), 에이콘, 2013, pp.318~321

27) Begas, BIG DATA, 빅데이터분석전문가 교재, 한국데이터베이스진흥원, 2016, p.343.

익률과 밀접하게 움직인다. 국고채 수익률이 상승(하락)하면 리츠수익률도 상승(하락)하는 것으로 나타난다. 이는 안전자산인 국고채 수익률이 상승하게 되면 국고채 수익률보다 리츠수익률이 높아야 하므로 리츠수익률 역시 함께 상승하기 때문이다.<sup>28)</sup>

이에 따라 본 연구에서는 자산운용사 및 판매사의 글로벌리츠재간접 투자현황을 데이터마트로 구성하여 국고채 5년 수익률의 최근 1년간 평균 수익률을 투가가치 기준으로 더미변수화한 최근 1년 수익률(IV)을 종속변수로 하였다.

독립변수는 금융투자협회에서 공시하고 있는 글로벌리츠재간접 자산운용사 및 판매사별 글로벌리츠투자 현황의 항목을 독립변수로 사용하였다.

세부 독립변수 항목으로는 금융투자협회의 글로벌리츠재간접 투자펀드별 실적데이터중 판매회사, 순자산총액, 보유자산 금액, 기준가격,

1개월 수익률, 6개월 수익률, 설정(전환)일 수익률, 운용보수, 판매보수, 판매수수료, 주요투자지역을 독립변수로 하여 투자성과를 예측하였다.

## 2) 기초통계분석

데이터마이닝을 활용한 투자성과 예측을 위한 분류분석과 군집분석을 위한 변수들의 빈도분석 결과는 아래 <표 5>와 같다.

데이터 마이닝을 활용한 분류모델과 군집분석 결과 변수들간의 상관관계분석(Pearson 상관분석) 결과 투가가치(IV)와 양(+의 상관관계)가 높은 독립변수는 기준가격(BP) 0.530, 1개월 수익률(OMR) 6.11이었으며, 음(-의 상관관계)가 높은 독립변수는 운용보수(OM) -0.610로 나타났다. 이러한 상관분석 결과를 통하여 1년 투자 수익률의 투자성과와 상관관계가 높은 변수는 기준가격, 1개월 수익률, 운용보수라 할 수 있다.

<표 4> 변수의 정의

변수명	판매회사	순자산 총액	보유자산 금액	기준가격	1개월 수익률	6개월 수익률	투가가치	설정(전환)일 수익률	운용보수	판매보수	판매 수수료	주요 투자지역
변수명 (영문)	Sales company	Net asset value	Amount of assets held	Base price	One-month return	Six-month return	Investment performance	Set (conversion) daily yield	Operation Maintenance	Sales rep	Sales commission	Major Investment Areas
변수명 (단축)	SC	NAV	AAH	BP	OMR	SMR	IV	SCD	OM	SR	SC	MIA
변수 설명	은행(1), 증권(2), 기타(보험, 투신등)(3)	단위:억원	단위:억원	1000이상(1) 1000미만(0)	%	%	* 투가가치: 최근(2016.10.31차 기준) 1년 수익률 성과(종속변수) 무위험수익률인 국공채3년물 최근 1년 평균수익률인 1.57%를 기준으로 투자성과 2.00% 이상은 투가가치 있음(1), 투자성과 2.00%미만은 투가가치 없음(0) * 설정, 운용보수, 판매보수, 판매수수료: % * 주요투자지역: 아시아(1), 글로벌(2), 일본(3), 아시아퍼시픽(4)					

<표 5> 빈도분석 결과

구분	SC_A	NAV	AAH	BP	OMR	SMR	SCD	OM	SR	SC	MIA	IV
평균	1.89	81.96	82.78	.09	-4.6154	-3.4776	13.1338	.78630	.823	.228	2.28	.24
표준편차	.638	106.107	106.819	.287	2.24492	2.09998	24.57379	.159400	.3416	.4198	.788	.430
분산	.407	11258.666	11410.307	.082	5.040	4.410	603.871	.025	.117	.176	.621	.185
최소값	1	1	1	0	-8.07	-6.74	-21.46	.400	0.0	0.0	1	0
최대값	3	376	379	1	-1.13	2.98	135.87	1.000	1.5	1.0	4	1

28) 조현민·김경민, "한국부동산시장과 자본시장과 부동산투자회사(REITs) 간의 연관성 분석", 국토연구, 국토연구원, 2011, 제71권, p.82.

### 3. 분석결과

#### 1) 의사결정나무(C5.0)

성과에 영향을 미치는 다양한 독립변수들을 분석한 결과 투자성과에 대한 의사결정나무 분석은 <그림 5>와 같다. 테스트데이터로 “투자가치 있음(1)”은 41건, 투자가치 없음(0)은 129건으로 총 170건의 데이터가 사용되어 분석되었다. 먼저 최초의 가지치기는 ‘1개월 리츠투자수익률(OMR)’을 기준으로 모형 검증결과 정분류율이 98.82%(트레이닝), 98.82%(테스트)로 높게 모형의 예측력이 확인되었다.

의사결정나무 모형을 통한 분석 결과 글로벌 리츠재간접 1개월 투자수익률(OMR) 투자성과를 예측하는 가장 중요한 변수로 작용한다는 것을 알 수 있었다.

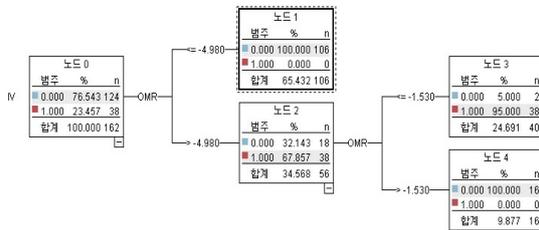
<표 6> 정분류율

'파티션'	1_훈련	2_테스트
정확	160 98.77%	168 98.82%
틀림	2 1.23%	2 1.18%
합계	162	170

\$C-IV의 일치 교차표(행은 실제 항목 표시)

'파티션' = 1_훈련		0	1
0		122	2
1		0	38
'파티션' = 2_테스트		0	1
0		127	0
1		2	41

<그림 3> 의사결정나무 노드



위의 결과를 종합해 보면 투자성과에 가장 큰 영향을 미치는 요인은 글로벌재간접투자의 1개월 투자수익률로 나타났다.

#### 2) 서포트벡터머신(SVM)

예측분석의 분류의 정확도는 <표 7>과 같다. 92.35%(트레이닝), 92.35%(테스트)로 높은 정분류율을 나타냈다. 따라서 SVM 모형을 활용하여 글로벌리츠재간접의 투자성과(IV)를 예측할 수 있는 설명력이 충분히 확보되었다고 할 수 있다.

아울러 SVM을 통하여 예측자의 중요도를 분석한 결과 투자성과(IV)에 영향을 미치는 중요도 분석 결과 OMR(1개월 투자수익률), MIA(투자지역), AAH(보유자산금액) 순으로 IV(투자가치)라는 종속변수 설명하는 중요한 변수임이 분석되었다.

<표 7> 정분류율

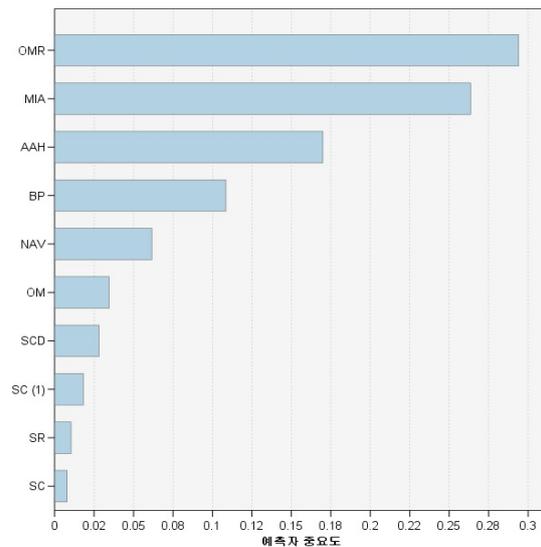
\$S-IV과(와) IV 비교

'파티션'	1_훈련	2_테스트
정확	155 95.68%	157 92.35%
틀림	7 4.32%	13 7.65%
합계	162	170

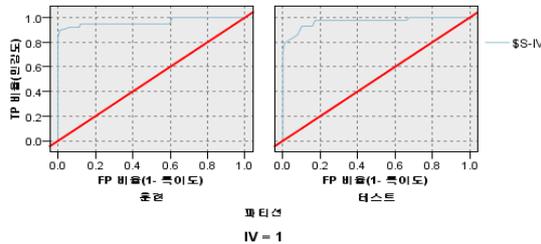
\$S-IV의 일치 교차표(행은 실제 항목 표시)

'파티션' = 1_훈련		0	1
0		123	1
1		6	32
'파티션' = 2_테스트		0	1
0		126	1
1		12	31

<그림 4> 예측자 중요도



<그림 5> ROC 그래프



강현철(2009)에 의하면 좋은 성능을 가진 모형 일수록 ROC 곡선은 대각선에 위치 한다. 이에 따라 SVM 예측분석을 통한 모형 판단은 좋은 성능 가졌다고 할 수 있다.

3) 인공신경망(ANN)

<표 8> 정분류율

\$N-IV과(와) IV 비교

정확	331	99.7%
틀림	1	0.3%
합계	332	

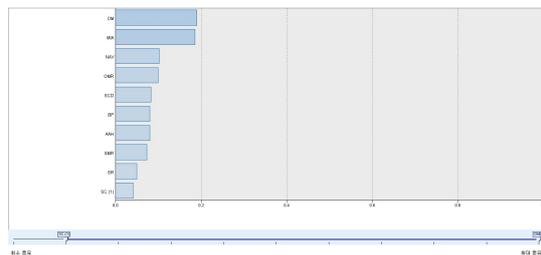
\$N-IV의 일치 교차표(행은 실제 항목 표시)

	0	1
0	250	1
1	0	81

성능 평가

0	0.28
1	1.398

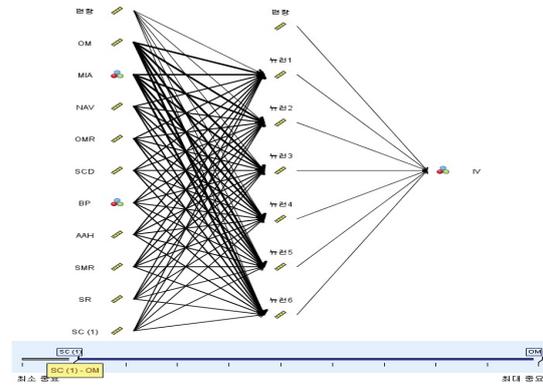
<그림 6> 예측변수의 중요도



예측분석의 분류의 정확도 위 <표 8>과 같이 99.7%로 높은 정분류율을 나타냈다.

인공신경망 모형(ANN)을 활용하여 글로벌리츠재간접의 투자가치(IV)를 예측할 수 있는 설명력이 충분히 확보 되었다고 할 수 있다.

<그림 7> 네트워크



인공신경망(ANN)을 통하여 예측자의 중요도를 분석한 결과 투자가치(IV)에 영향을 미치는 중요도는 OM(운용보수), MIA(주요투자지역), NAV(순자산총액) 순으로 중요도가 높은 것으로 분석되었다.

4) 로지스틱회귀분석(Logostic Regression)

전체의 분류의 정확도는 <표 9>와 같이 98.77%(훈련), 98.82%(테스트)로 비교적 높은 정분류율을 보였다. 이러한 예측의 정확성을 바탕으로 종속변수에 미치는 중요도는 BP(가격기준), NAV(순자산총액), OM(운용보수) 순으로 분석되었다.

<표 9> 분류표

'피타전'	1_훈련		2_테스트	
정확	160	98.77%	168	98.82%
틀림	2	1.23%	2	1.18%
합계	162		170	

\$C-IV의 일치 교차표(행은 실제 항목 표시)

'피타전' = 1\_훈련

	0	1
0	122	2
1	0	38

'피타전' = 2\_테스트

	0	1
0	127	0
1	2	41

성능 평가

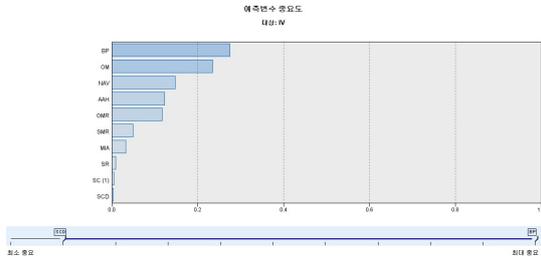
'피타전' = 1\_훈련

0	0.267
1	1.399

'피타전' = 2\_테스트

0	0.276
1	1.375

〈그림 8〉 예측변수의 중요도



## IV. 결 론

### 1. 연구의 결과 요약

데이터마이닝을 이용한 글로벌리츠재간접 펀드 투자 특성 예측 분석을 위한 기계학습 결과 신경망모형이 투자성과 예측을 위한 가장 높은 정분류율(99.7%의 정확도)을 보였다. 의사결정나무모형(C5.0)과 로지스틱회귀모형의 예측변수의 중요도 분석을 위한 예측의 정분류율(98.82%의 정확도)을 보였다.

예측의 정확도를 바탕으로 변수의 중요도 분석결과 종속변수인 투자성과(IV)에 영향을 미치는 독립변수는 OM(운용보수), 순자산총액(NAV), 가격기준(BP)이라는 사실을 확인할 수 있었다.

연구 성과는 부동산금융에도 데이터마이닝을 통한 빅데이터 예측분석을 통해 투자성과라는

투자성과의 특성을 기계학습 모델링을 통하여 실증분석 하였다는 점이다. 또한 연구모형을 실증에 거시지표 등 외부 요인이 아니라 실제 투자가 이루어진 투자모형내의 변수들을 통하여 확인하였다는 점이 연구의 큰 성과라 할 수 있다.

### 2. 연구의 시사점과 한계

리츠펀드는 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 글로벌리츠재간접 투자 운용담당자들에게 마케팅 분석시 기준을 제시하였다.

글로벌 리츠투자 세부 항목들 중 최근 1년의 투자수익률인 투자성과(IV)를 판단하는데 중요한 기준을 제시하였다는 점에서 시사하는 바가 크다 할 것이다.

본 연구는 데이터마이닝을 통한 예측분석과 중요변수 선정에 있어 독립변수를 다양하게 선정하여 예측의 설명력을 극대화 시키지 못하였다. 국내 리츠투자 수익률과의 단순 비교에 그치고 세밀한 계량화된 비교실증을 통해 글로벌리츠의 특성을 도출하지 못했다는 점은 연구의 한계라 할 것이다.

차재의 연구에서는 이점을 보완하여 국내외 리츠투자의 특성을 데이터마이닝을 통하여 예측하고 비교 분석하고자 한다.

### 參考文獻

- Tilo Wendler · Soren Grottrup, *Data Mining with SPSS Modeler*, SpringerVerlag, 2016.
- 이안위튼 · 아이베프랑크 · 마크홀, 데이터마이닝(기계학습의 이론과 응용), 에이콘, 2013.
- 캐빈머피, 머신러닝(Machine Learning), 에이콘, 2015.
- 방경식 · 장희순, 부동산학총론(제4판), 부연사, 2016.
- 김영곤외 4인, 부동산금융과 투자(제14판), 부연사, 2014.
- 박창이외 4인, R을 이용한 데이터마이닝, 교유사, 2015.
- 이재길, R프로그램에 기반한 다변량분석 및 데이터마이닝, 황소걸음아카데미, 2016.
- 여수진의 2인, "리츠 수익률의 결정요인에 관한 연구", 대한부동산학회지, 대한부동산학회, 2014, 제32권 제1호.
- 최차순, "REITs 수익률의 ARIMA Model 설정에 관한 연구", 부동산학보, 한국부동산학회, 2014, 제58권.
- 최효비의 1인, "은퇴계층의 부동산자산 운용에 관한 결정요인 분석", 부동산학보, 한국부동산학회, 2016, 제65권.
- 홍아름의 2인, "데이터마이닝 기법을 이용한 서울시 오피스빌딩 투자특성 예측에 관한 연구", 서울시연구, 서울연구원, 2010, 제11권 제2호.
- 전해정, "자산가격결정이론에 기반한 주택가격결정요인 분석에 관한 연구", 부동산학보, 한국부동산학회, 2013, 제52권.
- 이치주의 1인, "리츠와 건설경기, 부동산경기, 주식시장과의 관계분석", 한국건설관리학회지, 한국건설관리학회, 2010, 제11권 제5호.
- 조현민의 1인, "한국 부동산시장 및 자본시장과 부동산투자회사(REITs) 간의 연관성 분석", 국토연구, 국토연구원, 2011, 71권
- 김병준 · 이창석, "미국 REITs시장과 주식시장 수익률간의 상호 영향력 검증", 부동산학보, 한국부동산학회, 2016, 제66권
- 남연우 · 고석찬, "부동산펀드 성과 영향요인 분석", 부동산학보, 한국부동산학회, 2015, 제62권
- 김경민의 2인, "국민연금 기금운용을 통한 부동산투자 특성에 관한 연구", 부동산학보, 한국부동산학회, 2015, 제62권
- 조재영, "프로젝트 금융의 안정성 확보에 관한 연구", 부동산학보, 한국부동산학회, 2016, 제64권
- 이병훈의 3인, "회귀분석을 통한 부동산투자회사(REITs) 오피스빌딩의 수익률 영향요인 분석", 대한건축학회논문집, 대한건축학회, 2011, 제27권 제9호.
- 박덕배, "부동산 간접투자시장의 특징과 시사점", *VIP REPORT*, 현대경제연구원, 2015, 15-7호.
- 김홍년, "글로벌 리츠시장 현황 및 리츠산업의 활성화 필요성", 리츠저널, 한국리츠협회, 2015, VOL.16.
- 금융투자협회(<http://dis.kofia.or.kr/>)
- Fn-Guide(<http://www.fnguide.com/>)Index During The 2000-2011 Time Period", *Journal of Business & Economics Research*, 2014, 12(1).
- K.C..Tseng, "Forecasting Financial Stocks Using Data Mining", California State University Fresno, Forecasting, 2010.
- David. Enke · Suraphan Thawomwong, "The use of data Mining and neural networks for forecasting stock market returns", *Expert Systems with Application*, 2005, Vol.29.
- Christopher Grandrud, *Reproducible Research with R and RStudio*(second edition), CRC Press, 2015.