韓國IT서비스學會誌
第15卷 第4號
2016年 12月, pp.1-24

Journal of Information Technology Services
https://doi.org/10.9716/KITS.2016.15.4.001

비정형 데이터 분석을 통한 금융소비자 유형화 및 그에 따른 금융상품 추천 방법*

이재웅** · 김영식*** · 권오병****

Financial Instruments Recommendation based on Classification Financial Consumer by Text Mining Techniques*

Jaewoong Lee** · Young-Sik Kim*** · Ohbyung Kwon****

■ Abstract ■

With the innovation of information technology, non-face-to-face robo advisor with high accessibility and convenience is spreading. The current robot advisor recommends appropriate investment products after understanding the investment propensity based on the structured data entered directly or indirectly by individuals. However, it is an inconvenient and obtrusive way for financial consumers to inquire or input their own subjective propensity to invest. Hence, this study proposes a way to deduce the propensity to invest in unstructured data that customers voluntarily exposed during consultation or online. Since prediction performance based on unstructured document differs according to the characteristics of text, in this study, classification algorithm optimized for the characteristic of text left by financial consumers is selected by performing prediction performance evaluation of various learning discrimination algorithms and proposed an intelligent method that automatically recommends investment products. User tests were given to MBA students. After showing the recommended investment and list of investment products, satisfaction was asked. Financial consumers' satisfaction was measured by dividing them into investment propensity and recommendation goods. The results suggest that the users high satisfaction with investment products recommended by the method proposed in this paper. The results showed that it can be applies to non-face-to-face robo advisor.

Keyword : Fintech, Robo-Advisor, Text Mining, Decision Tree, Random Forest, Machine learning

Submitted : December 6, 2016 1st Revision : December 19, 2016 Accepted : December 23, 2016

* 본 연구는 정부(미래창조과학부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임(No. R0126-15-1007).

** 경희대학교 일반대학원 경영학과
*** 경희대학교 경영대학 경영연구원 연구원
**** 경희대학교 경영대학, 교신저자

1. 서 론

새로운 디지털 비즈니스 전략으로서 핀테크(Fintech)가 새롭고 가치 있는 재무금융서비스로 등장하고 있다. 핀테크는 전자결재부터 자산관리까지, 또한 개인간 금융거래부터 크라우드펀딩까지 아우르는 금융 산업을 재편할 수 있는 차세대 재무적 기술이며(Sia et al., 2016), 전통적인 재무금융서비스로부터 제공되는 재품 및 서비스를 혁신하는 재무금융서비스와 지능정보기술을 동시에 가진 기술 중심의 스타트업으로 정의되기도 한다(Chang et al., 2016).

재무금융서비스 고객들에 대한 데이터가 축적되면서 은행들은 새로운 고객 경험을 제공하고 데이터 분석에 기반한 새로운 가치를 창출하려고 한다. 이러한 새로운 가치 중 하나는 새로운 투자 고객의 개발이다. 전통적이고 고비용의 고객관리 방법을 보완하기 위해 Web2.0 또는 소셜미디어(Social Media) 기술을 활용하여 고객에게 가치가 있는 정보를 소비자들 스스로 창출하고 유통하게 함으로써 롱테일(long-tail)에 위치한 고객을 유지하거나 새로운 고객 개발 비용을 효율화한 것처럼, 핀테크, 특히 에이전트 기반의 디지털 어드바이저(digital advisor) 또는 로보 어드바이저(robo-advisor)도 기존의 고비용적 고객 투자 상담 외에 롱테일 군에 속하는 고객들에게 비용 효율적으로 고객의 투자성향을 파악하고 투자 상품을 추천하는 노력을 경감시킬 수 있다.

이를 위해 고객들이 자발적으로 남긴 데이터를 활용할 수 있을 것이다. 인터넷 상에 공유된 각종 투자 상품관련 오피니언이나 고객 상담 내용 중 VOC(Voice Of Customer)로 인식될 수 있는 자료들이 그 일부이다. 그러나 이들은 정형화되거나 투자의향을 파악할 목적으로 사전에 준비된 유형으로 준비되어 있지 않다.

더욱이 핀테크에서는 비대면 의사소통에 대응해야 한다. 따라서 직접 접촉하여 투자자의 성향을 파악하는 것이 거의 불가능하고 투자자가 남긴 음성 정보 혹은 텍스트 정보를 통해 간접적으로 투자자의 성향을 추론해야 한다.

금융상품의 복잡한 구조와 전문성, 시간성, 특수성 등의 금융상품 고유특성과 수익성에 대하여 금융소비자는 금융상품의 탐색과 이해 및 평가 등에서 어려움을 겪는다. 이러한 금융소비자의 텍스트로부터 추출할 수 있는 많은 정보 중에 유용한 정보의 하나가 금융소비자의 투자성향이다.

따라서 본 연구의 목적은 잠재적 또는 불특정 투자자가 공유한 온라인상의 비정형 텍스트를 분석하여 자동적으로 그에 맞는 투자 상품을 추천하여 주는 지능적 방법을 제안하는 것이다. 즉, 해당 문서의 주제에 대해 표현한 텍스트 문서를 작성한 금융소비자의 성향을 파악함으로써, 맞춤형 금융상품을 추천한다. 맞춤형 금융상품 추천이란 인터넷을 통하여 금융상품을 선택하려는 고객의 요구사항을 패턴으로 분석하여 고객에게 가장 적절한 금융상품을 실시간으로 설계 및 제공할 수 있는 자동화된 시스템이다. 궁극적으로는 개별 투자자의 금융소비 유형을 중시하는 요인에 따라 단어집합을 추출하고, 웹에서의 사용로그와 비교하여 사용자의 투자성향을 예측하고, 금융상품을 추천하려는 시스템을 제안하려고 한다.

본 논문은 다음과 같이 구성되었다. 제 2장에서는 현존하는 로보 어드바이저에 대해 정리하고 그 한계점과 개선 방안을 논의한다. 제 3장에서는 본 연구에서 제안하는 VOC와 같은 비정형 텍스트 및 란별 알고리즘 기반의 투자성향 추론 및 추천 방법에 대해 제안한다. 제 4장과 제 5장에서는 각각 제안 방법에 대한 성능을 검증하고 시사점과 결론을 제시하였다.

2. 로보 어드바이저

최근 로보 어드바이저(robo-advisor)로도 명명되는 자동 투자 거래(automated investment transactions) 또는 자동 투자 도구(automated investment tool)는 데이터 분석 및 지능형 기술을 활용

하여 잠재적 고객에게 직접 또는 투자상담사를 통하여 간접적으로 고객에게 유리한 맞춤형 투자 상품을 선택하도록 지원하는 지능형 도구이다. 로보 어드바이저라는 용어 자체는 최근 관심을 받기 시작했지만, 투자 거래를 전산 자원으로 지원한다는 기술로는 과거에는 자동 예약 및 발권, 전화기반 서비스와 같이 단일 혹은 주로 낮은 수준의 서비스나 현금지급기 형태로 존재해 왔다(Meuter et al., 2003; Dabholkar and Bagozzi, 2002; Hale and Thakur, 2006). 이에 비해 로보 어드바이저는 금융 투자 관련 유력한 자가서비스기술(Self-Service Technology, SST)로 기대되고 있다. 여기서 SST란 서비스 제공자가 생산성 또는 효율성을 제고할 목적으로(Walker et al., 2002; Zeithaml and Gilly, 1987), 고객들에게 그들이 편리한 채널을 통해 서비스에 스스로 접근하게 하여(Meuter et al., 2003) 고객들의 요구에 부합하고 만족도를 증진시키도록 하는(Bitner et al., 2002) 기술을 뜻한다(Liljander et al., 2006).

정보기술은 은행의 모든 측면에 포함되는 비즈니스전략에 요구되는 경영의사결정에 중요한 부분이다. 은행의 IT 전략은 비즈니스 전략과 융합되어 디지털 비즈니스 전략이라고 불린다(Bharadwaj et al., 2013). 현재 로보 어드바이저 업체와 같은 경우에는 IT기술을 활용한 자동화된 서비스로 낮은 수수료와 낮은 최소위탁자산을 주요한 IT기술과 친화적인 젊은 고객들을 대상으로 한 맞춤형 서비스의 제공하고 있다.

현재 존재하는 로보 어드바이저들은 다음 <Table 1>과 같다. 먼저 Fully-Automated Platform은 알고리즘 기반 시스템을 통해 개별 고객자산을 최적 자산배분하고 이에 따른 맞춤형 포트폴리오 자산운용을 Robo-advisor가 직접 하는 것을 말한다. Self-Executed Trades는 알고리즘 기반 시스템을 통해 고객 포트폴리오 모니터링 및 정기적 투자자문 제공하는 서비스를 말하며, Advisor-Executed Trades는 실제 어드바이저 판단으로 자문과 운용 업무를 수행하되 고객과의 커뮤니케이션 수단으로 온라인 채널을 활용하는 것을 말한다.

이로 보건대 성공적 활용을 위한 로보 어드바이저의 요건은 투자 고객의 성향을 잘 파악하고 그들 또는 그들과 같은 집단의 과거 투자 경험을 활용할 수 있어서 기술적으로 잘 준비된 기술이라는 점을 고객들에게 인식시키는 것이다. 그런데 아직 고객 또는 그들과 유사한 집단이 온라인상에 공유한 비정형 자료에 근거하여 투자 성향을 파악할 수 있고 이를 근거로 투자 상품을 자동으로 추천하는 로보 어드바이저는 거의 존재하지 않는다.

한편 자가서비스기술(SST)로서의 로보 어드바이저에 대한 성공적인 채택을 위해서는 다양한 요인이 필요하다. 문헌에 의하면 로보 어드바이저와

〈Table 1〉 Existing robot advisor

Robo-advisor	Organization	Description	Reference
Fully-Automated Platform	Wealth Front	Algorithm-based automated rebalancing, global investment portfolio using modern portfolio theory, Construction and operation	Wealth Front(2016)
	Betterment	Automated portfolio rebalancing, no limit on investment, free deposit and withdraw	Betterment(2016)
Self-Executed Trades	Market Riders	Algorithm-based portfolio monitoring management, Regular investment advice	Market Riders(2016)
Advisor-Executed Trades	Personal Capital	Based on modern portfolio theory, 6 Financial Asset Portfolio Management	Personal Capital(2016)
	Learn Vest	Financial management planning, budgeting, Online management service	Learn Vest(2016)

같은 STT의 성공적인 채택은 먼저 개인 특질에 의하여 결정될 수 있다(Dabholkar and Bagozzi, 2002; Susskind, 2004). 서비스 기술에 대한 통제를 고객 스스로가 할 수 있다고 느끼는 것은 매우 중요하며(Godin et al., 1992), 대기시간의 증가와 같은 사용하기 어려운 상황에서 SST를 사용하는 것이 도움이 된다는 확신이 영향을 미친다고 하였다(Meuter et al., 2000). 또한 SST의 기술적 준비도(technology readiness)와 같은 고객의 인지(Lin and Hsieh, 2006)나 유사 서비스 사용 경험(Dholakia et al., 2005), 그리고 SST를 사용하도록 하는 사회 환경적 요인(Reinders et al., 2008) 또한 SST의 채택에 영향을 주는 것으로 알려져 있다.

SST 관점에서의 로보 어드바이저의 성공은 투자성향 추론이나 금융서비스 추천에 있어서 개인의 특질, 고객 경험, 사회적 맥락(social context)에 대한 정보가 존재해야 그 정확도나 만족도가 제고된다는 점이다. 따라서 로보 어드바이저는 더 다양한 정보를 확보해야 하는데 기존과 같이 직접 사용자에게 질의하여 획득하는 방식은 불편(inconvenient)하고 거슬리는(obtrusive) 방법일 것이므로 개선이 필요하다.

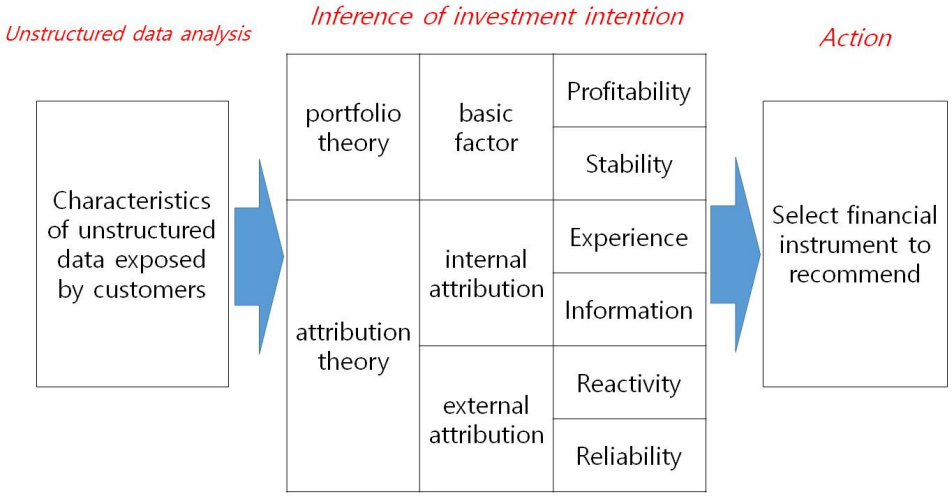
3. 방 법

3.1 투자성향 모형

로보 어드바이저 투자성향 추론 개선을 위해 <Figure 1>과 같이 고객의 투자성향에 대한 모형을 제안하였다. 본 모형은 크게 포트폴리오 이론과 귀인이론에 근거하였다. 즉, 고객이 온라인상에 노출한 비정형자료에서 추출한 단어적 특징을 통해서 투자의향을 파악한다. 투자의향은 귀인이론에서는 내적요인과 외적요인으로 나뉘며 내적요인으로는 경험성과 정보성, 외적요인으로는 반응성과 신뢰성, 그리고 포트폴리오이론에 근거할 때 수익성과 안정성의 어떤 성향인지를 추론한다. 즉 어떤 다 사람은 이상의 세요인 각각에서 어떤 유형(예: A는 경험성과 신뢰성 그리고 안정성에 속하는 자)에 속하게 된다.

3.2 금융소비자 투자성향

본 연구는 투자에 영향을 미칠 수 있는 금융소비자의 유형을 3가지로 분류하고 6가지 변수로 세분화하고자 한다. 앞서 살펴본 포트폴리오의 선택이론을 바탕으로 기본적 요인(수익성, 안정성)을



〈Figure 1〉 Investment Propensity Model

선정하였다. 그리고 귀인이론을 바탕으로 내적요인(경험성, 정보성), 외적요인(반응성, 신뢰성)으로 나누었다.

3.2.1 포트폴리오이론 : 기본 요인

투자를 하는데 있어 가장 기본적이고 중요한 함수는 개인의 기대수익과 위험이 될 것이다. 일반적으로 금융투자상품의 선택의 목적은 이익극대화 및 손실최소화에 있다(Zhou and Pham, 2004). 수익과 위험은 현대 투자론의 이론인 평균-분산 포트폴리오 분석이라는 투자행동의 기본적인 규범적 모델에서 지적하고 있듯이 투자대안을 선택하는데 있어 핵심적인 요인이다(Markowitz, 1952). 투자를 성공적으로 하기 위해서는 투자대상의 기대수익률과 위험을 동시에 고려하여 최적의 투자대안을 선택하는 것이 필요하다(Huh et al., 2010). 따라서, 투자에 대한 의사결정은 주로 위험-수익 간의 상충관계(risk-return trade off)가 있으며, 투자자가 기대수익과 위험의 관계를 어떻게 지각하고 이해하며, 선택 의사결정에 이용하는지 관찰하는 것이 핵심적인 개념이다(Ha and Kim, 2011).

위험은 일반적으로 투자대상이 가져다 주는 기대수익에 의한 불확실성을 의미한다. 투자자들은 전문지식이 아닌 일부의 지식만을 가지고 투자결정을 하기 때문에 투자행동은 위험 하에서 이뤄지는 대표적인 의사결정이다(Huh and Yoo, 2012). Ganzach(2000)에 따르면, 위험은 결과가 불확실한 대안들의 핵심적 특성으로 규정지으며, 사람들의 신기술, 채용, 재무 의사결정과 같은 분야에서 행동지침과 관련된 대안을 평가할 때 고려되어야 하는 가장 중요한 특징 중 하나로서 위험을 지적하고 있다. 위험이 투자안의 선택에 있어 중요한 이유는 위험을 인식하는 일이 수익추정과 함께 투자자의 의사결정에 지대한 영향을 미치기 때문이다.

투자안의 선택과 위험에 관련 선행연구들은 주로 부에 따른 위험추구성향이 어떻게 다르게 나타나는지 결론짓는 연구들이 다수가 관찰된다. Campbell (2006)은 부가 많을수록 위험회피성향이 낮고 반대로 부가 적을수록 위험회피성향이 높음을 보이며, 이는 가계의 투자활동에 큰 실수가 될 수 있음을 나타내었다. Warner and Cramer(1995)도 자산수준이 증가될수록 위험추구성향이 강하다고 주장한바 있다. Bertaut and StarrMcCluer(2000)은 소득 구분 상위 1%의 집단이 그렇지 않은 집단에 비해 더 높은 위험자산에 투자하고 있다고 실증결과를 보고하기도 하였다. 표본을 국내로 하여도 동일 한 결과를 제시하였다(Joo, 2008; Huh et al., 2010).

이 밖에 국내 기타연구로는 Han et al.(2000)에 따르면 수익과 손익상황에서는 위험회피성향이 나타내고 손실상황에서는 위험추구 성향이 강하게 나타난다고 주장한다. 그들의 연구에서는 위험의 행태를 시장세분화와 동시적으로 살펴보았다는데 의의가 있다. Min and Song(2014)의 연구에서는 위험 회피 성향을 측정하는 기법에 대하여 연구를 한 바 있으며, 문항을 통한 위험을 측정하는 것보다 확률 척도 방식을 더 좋은 방식이라고 주장한다.

3.2.2 귀인이론 : 내적 요인

경험성(experientiality)은 과거에 겪었던 투자경험이 현재 또는 미래의 투자의 의사결정에 영향을 미칠 수 있다는 특성이다(Andersen, 1993). 예를 들면, A사에 판매했던 펀드를 가입했던 투자자가 큰 손실을 경험했을 경우 향우 펀드라는 금융상품을 불신하여 펀드가입을 꺼려할 수도 있고, 이러한 투자자는 투자성향이 보다 위험회피형 투자자로 변모할 수 있다. 경험성은 객관화하기 어려운 척도이고, 주관적인 내적 척도로 구분할 수 있는 근거가 된다.

한편, 투자의 경험은 비록 나쁜 경험이라고 해도 나중에 투자시에 보다 주의 깊게 투자를 실행하기 때문에 꼭 부정적인 것으로 보아서는 안 된다. 투자의 경우 이러한 구매 의사결정은 구매경험에서 얻은 결과의 피드백으로 인하여 주기(cycle)를 형성하기 때문이다(Joo et al., 2009). 현재 시장의 투자결정은 과거 투자 경험을 통한 지식 의사결정에 의

존한다(Johanson and Vahlne, 1977).

과거의 투자경험은 금융자산을 배분하는 주요한 변인이며(Min and Song, 2014), 과거의 금융거래의 경험은 금융이해력을 높이는 계기를 만들기도 한다(Choi and Cho, 2011). Ban(2013)의 경우 2011년 펀드투자자조사 데이터를 이용하여 분석을 실시하였다. 만 25~64세의 수도권 및 지방광역시에 거주하는 2,576명을 대상으로 한 설문데이터를 분석하였는데 펀드투자경험은 공격적 투자성향, 높은 이익실현수준과 손실 감내수준을 보였다고 보고하였다. Joo et al.(2009)도 한국투자자교육재단에서 조사한 '제1차 펀드투자자조사' 자료를 이용하여 투자자의 펀드투자 의사결정과정에 따른 펀드투자 행태를 살펴보았는데 국내 펀드투자자는 수익률향상과 투자대상의 다양화를 위하여 펀드투자를 시작하게 되는데 투자의 주된 목적으로는 여유자금의 운용과 은퇴자금의 마련임을 보고하였다.

정보성(informativeness)은 금융회사가 제공하는 정보가 금융소비자라고 볼 수 있는 투자자들이 이용 가능한 정보를 제공하는 정도로 정의한다. 투자자로 통칭되는 금융소비자는 넓게 금융거래에서 금융업자와 금융계약을 체결하는 상대방으로 정의되고 있는데(Hwang, 2010), 금융시장에서 정보의 비대칭성이 존재하고 거래상대방과 충분한 교섭력을 가지지 못하기 때문에 정보성은 또 다른 측면에서 금융소비자의 보호와 관련이 있다(Joo, 2011). 그렇다면 금융거래에서 정보 비대칭이 생기는 근본적인 이유는 무엇인가? 이에 대하여 기존연구들의 주장을 정리하면 다음과 같다.

먼저, 상품의 특성에 영향을 주기 때문이라는 주장이다. 금융상품의 특성상 상품의 속성과 서비스의 속성이 결합된 특성 때문에 투자의 의사결정에 주요한 요인으로 작용할 수 있으며, 금융소비자의 정보 탐색행동이 일반적인 소비자의 정보탐색 행동과 차별적인 특성 때문이다(Choi et al., 2016). 이는 금융상품이 무형성과도 관련 있다. 형태가 없기 때문에 경험할 수 없으며, 주어진 정보에 의해서만 상품을 판단하고 결정해야 하기 때문이다.

또 다른 특성으로는 금융상품은 상당히 복잡하여 설명이 난해한 특성도 정보의 비대칭적인 성향을 부각시킨다. 다시 말하면, 전문성으로 요약할 수 있을 것이다. 기초지식이 있는 경우는 정보에 우위에 있다면, 투자의 실행이 용이하지만, 금융에 대한 기초지식이 없다면, 사실상 감(feeling)이나 경험에 의존한 투자를 실행할 수밖에 없다. Joo(2011)의 실험에서 50세 이상 금융소비자는 펀드투자시 금융기관 및 판매 직원에게 얻은 정보 및 의사결정을 의존하는 경향이 높고 펀드투자에 관한 이해력 수준이 50세 미만과 비교할 때 낮음을 알 수 있었다. 이는 중고령층이 취약한 금융소비자로 전락할 수 있음을 실증적으로 보였다는데 의의가 있다. Mishra and Kumar(2011)의 문헌에서도 확인할 수 있는 것처럼 뮤추얼 투자자의 지식수준이 높을수록 정보탐색행동이 활발하기 때문에 더욱 투자활동에 활발하게 됨을 감안할 때, 전문성의 확보는 정보 비대칭성이 감소함을 의미한다.

3.2.3 귀인이론 : 외적 요인

외적요인이라는 외부적인 요소에 의해 투자결정에 미칠 수 있는 요인을 말한다. 여기서 외부적인 요소라는 것은 투자자의 전체에 영향을 미칠 수 있는 요인을 말한다. 예를 들면, 거시경제 상황이나, 판매회사의 평판, 법 제도와 규제 등이 그 예일 것이다. 본 연구에서는 다양한 외적 요인 가운데 거시경제변수 등과 같은 거시적인 경제요인을 중요시하며 반응하는 반응성과 자산을 운용하는 운용회사나 판매회사에 대한 일반적인 평판에 대한 신뢰성을 변수로 선정하였다.

먼저 거시경제변수들은 투자심리에 영향을 미치는 주요한 변수이다. 일반적으로 재무론에서는 투자전략 중에서 반대투자전략과 계속투자전략이 있는데 반대투자전략은 주가수익률이 낮은 주식을 매입하고 주가가 높은 것을 매도하는 행위로 이는 거시경제변수가 영향을 미치지 않음을 시사한다. 반대로 계속투자전략(주가가 높은 것을 매입하고 낮은 것을 매도)이 성립할 경우 거시경제변수가 영향

을 미치고 있음을 의미한다. Jegadeesh and Titman (1993)은 계속투자전략이 유리하고, Chordia and Shivakumar(2002)의 경우에는 과거 거시경제변수를 통해 살펴볼 때 계속투자전략이 유효한 전략임을 주장한다. 반면, Bondt and Thaler(1985), Chan (1988)는 하락하는 시장에서의 하방위험(downside risk)에 대해 투자자들은 보다 큰 관심을 가지기 때문에 비록, 펀드가 좋지 않은 성과를 낼지라도 펀드 투자자의 펀드에 대한 수요가 증가한다고 주장한다. 국내 자료를 분석한 Lee et al.(2014)도 불확실성이 증가하는 시기에 국내 주식형펀드 투자자들은 주식형펀드로의 자금 유입을 증가시킨다고 주장한다. 이처럼 거시경제변수의 변화에 따라 주가는 오를 수도 있고 내릴 수도 있다.

여기서 중요한 사실은 거시경제변수가 직접적으로 자산 가격을 상승하거나 하락시키는 방향성이 중요한 것이 아니라 상승과 하락의 변동성을 가지게 된다는 점이다. 즉, 수익률의 불확실성을 갖기 때문에 투자심리도 변화할 수 있다는 것을 유추할 수 있다. 따라서, 거시 경제 변수의 움직임은 투자를 실행하는데 있어 중요한 요인임을 시사한다. 반응성은 이러한 불확실성 하에 나타나는 통제 불가능한 거시경제변화에 따른 금융소비자의 반응을 의미한다.

본 연구에서 중요하게 생각하는 외적요인의 다른 요소로 신뢰성에 주목한다. 신뢰성은 금융상품을 판매하거나 운용하는 회사의 명성이나 이미지, 무형자산 등을 투자자가 신뢰할 만한지에 대한 객관적인 것을 의미한다. 운용하는 기업의 명성이 높은 경우, 기업의 경영활동에서 실질적으로 나타나는 주가상승과 매출증가 등의 경제적인 이득뿐만 아니라 경제적인 수치로 환산하기는 어렵지만 기업의 생존에 장기적인 이득을 가져다준다(Lyon and Cameron, 2004). 예를 들어, 대중들은 평소 높은 명성을 가진 기업의 부정적인 신문보도를 접할 경우, 그렇지 않은 기업들의 부정적인 신문보도를 접할 경우보다 덜 부정적인 태도를 형성하는 것으로 나타났다(Cha et al., 2007).

금융산업은 서비스산업 내에서도 비교적 고객의 중요성을 먼저 인식하고 고객서비스를 산업의 핵심 역량으로 보며, 고객만족을 경영의 최고 가치로 인정하고 있다(Lee, 2012). 또한 고객만족을 위한 서비스 품질은 경쟁이 더욱 치열해지고, 은행이 생존의 어려움에 처해 있을 때 한층 더 관심을 가지고 신경을 써야만 하는 요인이다(Lee, 2012). 이는 대중이 상대적으로 높은 명성 가진 기업을 더 신뢰하며 정직하다고 생각하고 선호하기 때문이다(Lyon and Cameron, 2004). 금융마케팅의 고전인 Anderson(1982)은 금융소비자가 점포를 선택하는데 있어 본점의 평판도 매우 중요한 요인임을 실증결과로 나타냈다. 따라서, 유형적인 상품뿐만 아니라 무형적인 서비스 등이 소비자가 느끼는 금융기업의 명성에 많은 영향을 끼치고 있으며 금융권에서의 무형적인 서비스와 관련된 소비자의 심리적 속성들이 서비스 모델을 개발하는 데 활발하게 이용되고 있다(Hur and Kim, 2001). 이처럼 금융기관의 이미지나 평판을 의미하는 신뢰도가 투자의 감정과 행동에 영향을 미치고 있음이 선행연구들에 의하여 확인되고 있다.

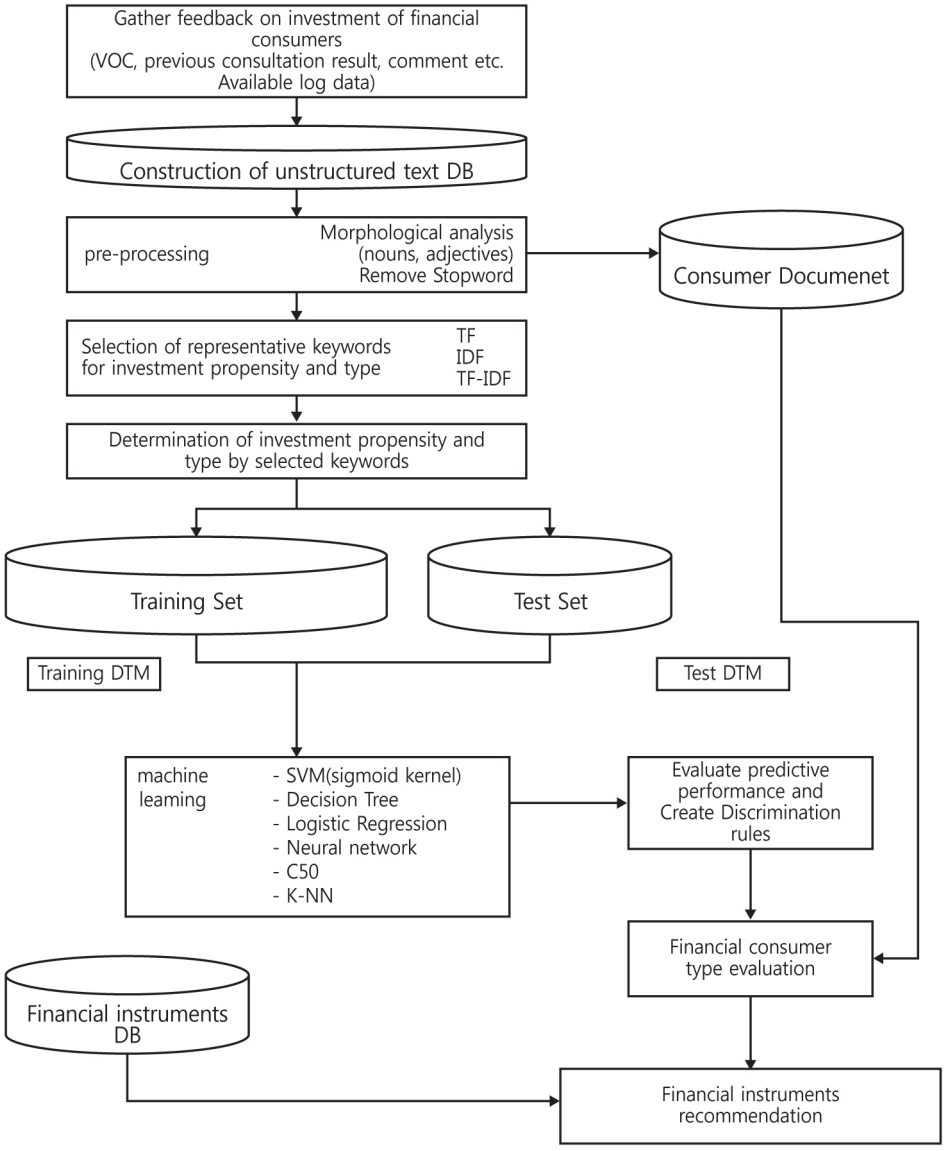
먼저 귀인이론(Attribution theory)은 인지과학이나 사회 심리학 등 사회과학 분야의 영향력 있는 이론으로 평가되어진다(Zeelenberg and Pieters, 2004). Heider(1958)는 인과에 관련된 지각의 핵심문제는 어떤 특정의 행위를 내부 상태들이나 외부 압력에 귀인시킬 것인가라고 하였으며 행위에 영향을 미치는 인적, 환경적 요소로 나타내어 설명하였다. 귀인이론은 소비자 의사결정을 하는데 있어 원인과 결과의 관계를 추론하는 경향이 있으며, 이것이 소비자의 감정과 소비자의 태도 그리고 소비자 행동에 영향을 미친다는 것을 나타낸다(Sung and Han, 2010).

3.3 전체적인 프로세스

금융소비자의 투자 유형을 해당 금융소비자가 투자에 대해서 남긴 글(VOC, 이전상담결과, 댓글

등 가용한 로그 데이터)을 기반으로 자동 파악하고, 그에 근거하여 금융 상품을 추천하는 방법을 <Figure 2>와 같은 프레임워크로 제안한다. 제안 프레임워크는 크게 세 단계로 나뉘어진다.

첫째로 금융소비자인 실험대상자가 투자에 대해서 남긴 글(VOC, 이전상담결과, 댓글 등 가용한 로그 데이터)을 수집하여 검증용 문장으로 삼는다. 둘째, 검증용 문장으로 말뭉치를 구성하고, 이에 대해 형태소 분석, stopword 제거, 명사 추출 등의 전처리 작업을 수행한다. 셋째, 추출된 단어들 각각의 통계치(TF, IDF, TF-IDF 등)를 계산하여 투자성향 및 투자 유형을 나타내는 대표 키워드를 선정한다. 선정된 키워드 별로 특정 투자성향 또는 유형인지를 판별하는 통계치를 구하여 <키워드, 통계치>



〈Figure 2〉 Overall Framework

벡터를 생성하고, 이 벡터 값을 학습데이터로 하여 다양한 판별 알고리즘을 수행함으로 키워드와 잠재된 투자 성향 사이의 관계를 판별하는 규칙을 생성한다. 예측 성능평가를 실시하여 성능분석결과 가장 우수한 것으로 선정된 판별 알고리즘을 활용하여 실험대상자의 투자성향을 추론한다 마지막으로, 추론한 투자성향으로 투자상품을 실험대상자에게 추천하고 그 정확성을 점검한다.

3.4 조사대상

본 연구는 사전조사를 통해 투자성향을 판단할 수 있는 설문을 개발하고, 모델의 성능 평가를 위해 20~60대의 투자의향 및 투자여력이 있는 800명의 성인 남녀를 대상으로 설문 데이터를 수집하였다. 설문지에서 텍스트정보를 얻을 수 있는 간접 설문과 설문자의 투자성향과 관련된 정보를 얻을 수 있는 직접 설문항목을 설정하여 설문을 실시하였다.

투자자들이 투자 성향을 어떻게 표현하는지에 대해 수집하기 위해 예를 들어 다음과 같은 형식의 질문을 하였다.

ㅇㅇ사 ㅇㅇ제품(채권)은 국내 우량채권에 투자하여 장기적으로 안정적인 수익을 추구하는 국내 최대 규모의 공모 채권형 펀드입니다. 투자에 관심 있으십니까?

그러면 응답자는 예, 아니오로 답한 다음 그 이유에 대해서 비교적 상세하게 그 이유를 작성하도록 하였다.

본 연구는 사전조사를 통해 투자성향을 판단할 수 있는 설문을 개발하고, 모델의 성능 평가를 위한 금융소비자 설문조사를 시행하였다. 소비자 패널이 구축되어 있는 업체 가운데 선정하였는데, ㈜마크로밀엠브레인이 조사를 대행하였다. 설문은 온라인 조사로 실시하였으며, 30세 이상의 투자여력이 있는 전국의 각 800명을 대상으로 2016년 6월 29일~2016년 7월 5일까지 7일간 수행되었다. 총 800명이 설문에 응답했지만 불성실 설문응답 8개를 제외한 792개 데이터를 최종적으로 분석대상으로 선택하였다. 유효한 설문응답자의 인구통계학적 특성은 다음 <Table 2>와 같다.

〈Table 2〉 Descriptive Statistics

Characte-ristics	Categories	Partici-pants	Distri-bution
Gender	Men	385	48.6%
	Women	407	51.4%
Age	30~39	259	32.7%
	40~49	264	33.3%
	50~59	217	27.4%
	Above 60	52	6.6%
Level of education	High school and below	109	13.8%
	College	100	12.6%
	Bachelor	466	58.8%
	Master and above	117	14.8%
Job	Agriculture/Forestry/ fishing	2	0.3%
	public official	17	2.1%
	Education	23	2.9%
	Profession	60	7.6%
	Business management	63	8.0%
	office job	277	35.0%
	Production/Technical	50	6.3%
	Service/Sales	35	4.4%
	Self-employment	58	7.3%
	Freelancer	32	4.0%
	House wife	149	18.8%
	Student	1	0.1%
	inoccupation	17	2.1%
	ETC	8	1.0%
disposable income	below ₩1,000,000	72	9.1%
	₩1,000,000~₩5,000,000	268	33.8%
	₩3,000,000~₩5,000,000	258	32.6%
	₩3,000,000~₩7,00,000	117	14.8%
	above ₩7,000,000	77	9.7%
Total (N)		792	

설문조사 표본의 일반적 특성으로는 <Table 2>와 같으며, 연구에서는 금융소비자 특성 중에서 성별, 연령, 학력 등의 인구통계학적 특성을 파악하기 위하여 빈도분석을 실시한 결과, 성별에 있어서 792명의 응답자 중, 385명이 남성, 407명이 여성으로 고르게 분포하였다. 또한 연령대도 투자 능력이 떨어지는 30세 이하는 제외하였으며 30대, 40대, 50대 이하를 고르게 수집하였다. 직업은 대부분 직장을 가진 자를 대상으로 하였으며, 전업주부도 가처분소득이 있는 자는 대상에 포함하였다.

3.5 말뭉치 구성 및 전처리

수집된 텍스트 데이터로부터 단어를 추출하기 위해 다음과 같은 전처리 과정을 수행하였다. 먼저 연구소에서 자체 개발한 RHINO 3.4.4을 이용하여 설문 응답자의 특정 투자 상품에 대한 평가 텍스트를 형태소 분석하였다. <Table 3>은 수집된 금융소비자의 데이터를 명사 및 형용사를 추출한 양식으로 텍스트 파일로 저장된 데이터를 'R'을 이용하여 csv파일로 정리한 양식이다. 그리고 구글에서 제공하는 stopword 리스트를 활용하여 의미 없는 단어들을 제거하였다. 여기서 유형이란 설문 응답자가 스스로를 정의한 것으로 나중에 투자성향 분류 학습을 위한 데이터로 활용된다.

3.6 투자성향별 대표 키워드 선정

형태소 분석을 통해 추출된 명사들은 투자성향을 나타내는 정보를 구성한다. 문서 분류과정에서 일반적으로 사용되는 문서처리는 학습과정과 분류과정에서 분류기가 인식할 수 있는 데이터로 변환한

〈Table 3〉 Noun Extraction Using RHINO 3.4.4

Consumer type	Extract noun
Profitability	투자처, 주식, 대상, 금액, 증권, 이상, 단기, 수익, 위험, 기회, 돈, 리스크, 이익, 최대, 메리트, 개인, 뒤, 수, 좋, 높 등(investment, stock, target, amount, securities, ideal, short, profit, risk, opportunity, money, risk, profit, maximum, merit, personal, back, number, good, high, etc.)
⋮	⋮
Stability	기초, 은행, 회사, 믿음, 요건, 안정, 수익, 안전, 채권, 우량주, 블루, 성향, 연, 연간, 중도, 대표, 안, 범위, 만큼, 어렵, 때문 등(base, bank, company, faith, requirement, stability, profit, safety, bonds, blue chip, blue, orientation, annual, middle, representative, because, etc.)
⋮	⋮
Information	운용, 분야, 형태, 필요, 설명, 관심, 이유, 연금, 수수료, 관심, 율, 연말, 환매, 부담, 경고, 지역, 이자, 안내, 뿐, 것, 외, 없 등(operation, Field, form, need, explanation, interest, reason, annuity, fee, interest rate, year-end, repurchase, burden, warning, area, interest, guide, only, other, no, etc.)
⋮	⋮
Experience	펀드, 주식, 국외, 관련, 코스닥, 개인연금, 국내, 지금, 코스피, 가치, 성장, 때, 함, 경기, 수익, 연금, 자체, 손실, 만회, 때문, 힘들, 같 등(funds, stocks, foreign, related, KOSDAQ, personal pension, domestic, now, kospe, value, growth, time, game, profit, pension, self, loss, retracement, difficulties, same, etc.)
⋮	⋮
Reliability	장기, 언론, 회사, 신탁, 평가, 성장, 기준, 기관, 계열, 판단, 전문, 수익, 기업, 투자, 할, 높, 어떻하, 무엇 등(long term, media, company, trust, evaluation, growth, standards, agency, judgment, expertise, profit, enterprise, invest, high, how, what, etc.)
⋮	⋮
Reactivity	안정, 현시점, 국내, 채권, 은행, 우량, 금리, 변동성, 침체, 세계, 시트, 글로벌, 블랙, 보도, 없, 수, 어렵 등(stable, current, domestic, bond, bank, superior, interest, volatility, stagnation, world, sheet, global, black, sidewalk, none, water, difficult, etc.)

다. 792개의 문서에서 1,024개의 단어들로 구성된 Document-Term Matrix로 만들어진 TF-IDF가중치는 <Table 4>와 같은 양식으로 정리되었다.

비정형 텍스트 문서를 정형화하여 표현하기 위해 단순히 한 문서에서 특정 단어의 출현빈도(Term Frequency : TF)를 이용하여 표현하는 방법과 특정 단어의 출현빈도 정보에 특정 단어가 전체 문서에서 얼마나 희소하게 등장하는지를 의미하는 역문서빈도(Inverse Document Frequency : IDF)의 정보를 함께 사용하여 가중치를 표현하는 두 가지 방법으로 구분된다. 본 연구에서는 TF-IDF 가중치를 이용하여 단어집합의 중요도에 따라 가중치를 부여하여 분류기가 인식 할 수 있는 데이터로 변환한 문서의 벡터 수치화 방법을 선택하였다. 이를 통하여 얻은 TF-IDF 가중치 값을 이용하여 Document-Term Matrix를 구성하여 나타내었다.

3.7 말뭉치 구성 및 전처리

다음으로 금융소비자가 남긴 텍스트를 근간으로 그의 투자성향을 파악하기 위한 판별 규칙을 생성하기 위해 여러 판별 알고리즘을 수행 비교하였다. 이때 사용한 판별 알고리즘은 SVM(sigmoid kernel), Decision Tree(CART), Logistic Regression, C50, k-NN, Random Forest이며 데이터마이닝, 통계, 데이터베이스 저장 기술, 패턴 인식, 정보 수집, 웹 검색, 마케팅, 생물학을 포함하여 다양한 분야에서 판별 알고리즘이 사용되고 있다.(Song and Kim, 2012; Wu et al., 2003). 본 연구에서는 금융소비자가 Web에 남긴 텍스트 데이터를 통하여 투자성향을 파악하는데, 텍스트는 unstructured document로부터 표면적으로 드러나지 않는 흥미로운 사실들을 발견할 수 있으며(Griffiths and Steyvers, 2004), 실시간으로 등장하는 많은 양의 데이터를 객관적이고 정확하게 처리할 수 있다. 최근 데이터베이스에 저장되는 대용량의 빅데이터 시대가 도래함에 따라 대용량의 텍스트에 대한 마이닝 연구가 활발하게 이뤄지고 있다.

먼저 서포트 벡터 머신(support vector machine, SVM)은 서로 다른 클래스에 속한 데이터들 간의 간격이 최대가 되는 초평면을 찾아 이를 기준으로 데이터들을 분류하는 알고리즘이다(Cortes, 1995). 설문과 같은 주관적이고 감정적인 표현을 포함한 텍스트 데이터를 SVM과 같은 기계 학습 알고리즘에 의해 학습하고 새로운 문서를 자동적으로 분류하는 연구들은 활발하게 진행되고 있다(Manevitz and Yousef, 2001; Joachims, 1998). 각 데이터간의 유사성을 계산하고, 구조적 리스크 최소화원칙(structural risk minimization : SRM)을 이용하여 일반화 오류를 줄이는 방식으로 유사 영역의 데이터 그룹들로 분류한다. 때문에 패턴인식과 문서범주화 등에서 우수한 성능을 나타낸다. 이미 여러 학자들에 의하여 그 우수성이 증명된 방법이다

〈Table 4〉 Document-Term Matrix

No	Interest	Conclusion	Product	Pension	Recession	Investment	Effect	Consumer type
1	0	0	0.455828	0.518526	0	0	0	stability
2	0	0	0	0	0	0.101465	0.942071	stability
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
117	0	0	0	0	0	0	0	information
118	0	0	0	0.129631	0	0.112738	0	information
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
352	0.158942	0	0.290072	0	0.507596	0	0	reactivity
353	0	0	0	0	0	0.059685	0	reactivity
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

(Burges, 1998).

둘째, 기존의 회계학에서 널리 사용되어온 회귀식을 판별규칙을 생성하는 데 이용하는 Logistic Regression 분석 방법을 이용하였다. 로지스틱 회귀는 Cox(1958)가 제안한 확률모델로, 독립변수의 선형 결합을 이용하여 종속변수와 독립변수간의 관계를 함수로 나타내고, 특정 관측치가 종속변수에 속할 확률을 알려준다. 본래 로지스틱 회귀분석은 이항 분류문제에 사용되나 본 연구는 3개 이상의 '1 of m' 분류 문제이므로 multinomial logistic regression을 사용하였다. 로지스틱 회귀분석은 기업도산예측, 주가지수 예측, 금융사기탐지, 고객구매 예측모형 등 다양한 경영의사결정분야에 널리 적용되어 왔다(Barniv, 1997; Zavgren, 1985; Bell et al., 1990; Ngai, 2010)

셋째, Decision tree의 대표적인 알고리즘으로 속성 선택 방법에 따라 지니계수를 사용하는 CART 그리고 C4.5(Quinlan, 1996) 등이 있다. 본 연구에서는 C4.5에서 발전시킨 C5.0(Quinlan, 1997)을 이용하였다. 데이터의 특징이나 속성에 따라 Tree형태의 의사결정 훈련 데이터모델을 만들고, 반복을 통해 주어진 문제에 대한 최종 결정에 따라 Tree 구조의 결과로 표현된다. 데이터 패턴이 간결하고 쉽게 이해하고 설명할 수 있는 장점이 있어, 분류와 예측을 하는 데 있어 효과적으로 사용된다.

다음으로 인스턴스 기반인 k-NN(k-nearest neighbor) 알고리즘은 1950년대 초에 첫 등장했다. 대용량의 training set이 주어졌을 때, 좋은 성능을 나타내는 기법이지만, 컴퓨터가 발전 보급되는 1960년대까지는 주목 받지 못했다. n개의 속성을 가진 training data를 n차원의 공간에 투사하고 test data가 어느 training data와 비슷한지 비교하는 방식으로 분류작업을 수행한다. 즉 n차원의 공간상의 training data들 중 test data와 가장 가까운 k개의 이웃(neighbor) 찾는다(Tan, 2006). test data를 입력하기 전까지는 training data를 인스턴스로 저장해두고 모델의 생성을 미뤄두기 때문에 인스턴스 기반의 학습자(instance-based learner)라고도 부르며 lazy learner라고도 부른다(Jiang et al., 2012).

마지막으로 Random Forest을 분류 알고리즘으로 선정하였다. Random Forest의 경우, Breiman (2001)에 의해 개발된 분류기법으로, 전통적인 Decision Tree 기법인 아닌 복수개의 나무로 Decision Tree를 구성하고 클래스를 개별 트리로 출력하는 앙상블 분류 기법으로 메타학습 형태를 갖고 있다. 새로운 개체의 클래스를 분류하기 위하여, 각각의 트리는 하나의 분류 클래스를 제공하며, 이들 트리의 다수결 투표에 의하여 최종 클래스가 결정되는 방식으로 진행된다.

이처럼 다양한 기계학습 기법들이 제안되고 적용되어 왔지만 분류기법 적용에 대한 각 분야별 텍스트의 성격이 많은 차이를 보이고 있기 때문에, 여러 알고리즘들에 대한 학습을 통해, 가장 적합한 분석 알고리즘을 활용하고자 한다.

본 연구는 전문가 설문을 통해, 설문자에게 받은 Text 데이터에 대하여 전문가가 귀인이론 중 내적요인(경험성, 정보성)과 외적요인(신뢰성, 반응성) 그리고 포트폴리오의 선택이론을 바탕으로 기본적 요인(수익성, 안정성)으로 판단하였고 그 결과, 두 차례에 걸친 coding을 통해 분류를 진행한 가중치 값에 대해 라벨링을 실시하였다.

4. 실 험

4.1 실험설계

사전조사를 통해 투자성향을 판단할 수 있는 설문을 개발하고, 모델의 성능 평가를 위해 30~60대의 투자의향 및 투자여력이 있는 총 800명의 설문자 가운데, 불성실 설문응답 8개를 제외한 792명의 성인 남녀를 대상으로 설문 데이터를 수집하였다. 이 문서를 토대로 특징적 단어 집합을 구성하였다. 한글문서에서 명사를 추출하기 위해 RHINO 3.4.4을 사용하였고, 각 단어 집합을 TF-IDF가중치로 변환하여 구성하였다. 이 중 20%에 해당하는 텍

스트데이터는 예측정확성 측정을 위해 예측데이터로 따로 저장을 하였다. 트레이닝 셋의 데이터가 테스트 셋에 영향을 미치지 않게 하기 위하여 홀드아웃 방법(hold-out)을 사용하여 수집된 데이터를 분류기 학습을 위한 트레이닝 셋과 예측 성능 평가를 위한 테스트 셋으로 나누고 R함수를 활용하여 매 반복실험마다 70대 30의 비율로 데이터를 무작위로 나누었다. SVM(sigmoid kernel), Decision Tree (CART), Logistic Regression, Decision Trees and Rule-Based Models(C50), Knn, Random Forest 알고리즘을 이용하여 분류를 위한 학습을 실시하였다.

학습알고리즘의 성능 비교를 위해 Confusion matrix에서 도출되는 accuracy, precision, recall 및 F-score를 metric으로 선정하였다.

4.2 결과

본 연구에서는 투자성향에 대한 정보를 설문을 통해 얻은 후, 학습하는 방법으로 기계학습 알고리즘을 응용하여 사용자의 거래 성향을 분석하고 실험을 진행하였다. 분류기로 트레이닝 셋을 학습시킨 다음, 테스트 셋을 투입하여 그 성능이 어느 정도 인지를 평가하였다.

<Table 5>~<Table 7>은 문서 가중치를 이용하여 hold-out방식으로 100회를 실험하여 얻은 결과의 평균값이다. 학습된 분류기의 성능을 정확하게 평가하기 위해 전체 데이터 셋의 분할부터 분류기의 학습까지의 단계를 100회 수행한다. 문서의 분류 성능을 평가하기 위한 기준은 accuracy, precision, recall, F-score 측정 식을 사용하였다.

〈Table 5〉 Performance Evaluation of Classification Based on the Attribution theory(internal attribution)

	accuracy	Precision	Recall	F-score
SVM(sigmoid kernel)	0.6122	0.5153	0.5220	0.5186
Decision Tree(CART)	0.6905	0.5339	0.6504	0.5856
Logistic Regression	0.6090	0.5347	0.5330	0.5338
Decision Trees and Rule-Based Models (C50)	0.6365	0.5098	0.5335	0.5211
k-NN	0.5661	0.3994	0.4806	0.4357
RandomForest	0.7153	0.5324	0.7727	0.6378

〈Table 6〉 Performance Evaluation of Classification Based on the Attribution theory(external attribution)

	accuracy	Precision	Recall	F-score
SVM(sigmoid kernel)	0.7229	0.5703	0.5750	0.5723
Decision Tree(CART)	0.7868	0.6015	0.6847	0.6396
Logistic Regression	0.7420	0.5991	0.6036	0.6009
Decision Trees and Rule-Based Models (C50)	0.7538	0.5238	0.6190	0.5665
k-NN	0.7340	0.4481	0.5834	0.5055
RandomForest	0.7837	0.5129	0.7384	0.6040

〈Table 7〉 Performance Evaluation of Classification Based on the Portfolio theory

	accuracy	Precision	Recall	F-score
SVM(sigmoid kernel)	0.4969	0.4823	0.4798	0.4810
Decision Tree(CART)	0.5190	0.4943	0.5132	0.5031
Logistic Regression	0.4776	0.4649	0.4648	0.4649
Decision Trees and Rule-Based Models (C50)	0.5012	0.4743	0.4788	0.4765
k-NN	0.3689	0.3862	0.4019	0.3938
RandomForest	0.5961	0.5497	0.5926	0.5703

먼저 귀인이론을 바탕으로 내적요인(경험성, 정보성, 이외)에 의한 분류를 바탕으로 분류알고리즘의 성능을 평가한 결과, Random Forest의 F-score가 약 63.78%로 가장 좋은 성능을 보였으며 Decision Tree(CART)가 58.56%, Logistic Regression이 53.38%로 그 뒤를 이었다.

또한 외적요인(반응성, 신뢰성, 이외)을 바탕으로 한 분류 알고리즘 성능 비교의 경우, CART방법의 F-score가 63.95%로서 가장 좋은 예측율을 보였고, 다음으로 RandomForest가 60.4%로 높은 성능을 보였다.

한편 포트폴리오이론을 바탕으로 한 기본요인(수익성, 안정성, 이외)을 분류한 알고리즘 성능 비교의 경우, RandomForest의 F-score가 57.03%로서 가장 좋은 예측율을 보였고, 다음으로 decision Tree (CART)가 50.31%로 높은 성능을 보였다.

결론적으로 전체적인 평균 예측력을 비교해보면 Random Forest, Decision Tree(CART), Logistic Regression, C50, SVM(sigmoid kernel), k-NN순으로 상대 모형들에 비해 최소 3%에서 최대 20%까지 성능의 차이가 나타남을 알 수 있었다. 특히 k-NN분류자의 경우 비교대상 판별 알고리즘들 중 가장 낮은 성능을 보였는데, text classification의 경우 high dimensionality의 문제가 존재하는데 k-NN의 경우 high dimensionality에 민감하기 때문으로 보인다. 각 data set의 성능의 평가정보를 비교해 보았을 때, Random Forest 방법이 가장 유의하게 나타났음을 확인하였다.

4.3 투자상품 추천

제 4.2절의 결과에 따라 소비자유형화에 가장 적합한 알고리즘으로 Random Forest 방법을 채택하여 금융소비자 유형에 따른 투자상품 추천을 제안하도록 했다. <Table 8>은 본 연구에서 사용한 분류 성능측정을 위한 금융소비자 데이터의 예시이다.

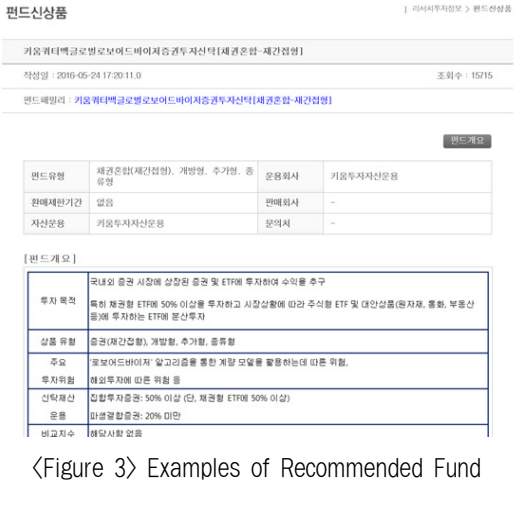
본 연구에서는 각 투자상품의 성격과 수익률 그리고 운용규모에 따라 추천펀드를 선정하였고, Random Forest의 결과를 채택하여, 금융소비자의 유형과의 매칭하도록 하였다. 펀드에 관련한 자료는 펀드 평가 전문 기관인 (주)제로인의 펀드추천(www.funddoctor.co.kr) 사이트로부터 금융소비자의 글을 통해 분류된 투자유형과 URL을 제시하였다. <Figure 3>은 펀드추천 콘텐트의 예이다.

〈Table 8〉 Measurement Data of Financial Consumer

1	oooo에 관심은 있지만 고배당주가 뭔지는 모르겠다 투자에 큰수익을 남기고 안전할 수 있으면 해보고 싶다 그런 게 없겠지만.(I'm interested in 0000 investment trust but I don't know what a high dividend. I want to make a big profit from investment, if I can be safe.)
2	왠지 복잡할 것 같은 생각이 듭니다. 그리고 개인연금 펀드는 왠지 신뢰성이 안갑니다. 그냥 일반 주식투자가 더 나을 것 같습니다.(I think it's kind of complicated. And private pension funds are somehow unreliable. It just seems like a better investment in stocks.)
⋮	⋮
86	위험부담은 있으나 고배당에 관심이 많아서, 투자기간이 어떻게 되는지 궁금합니다, 최소투자금액이 얼마나 되는지 궁금합니다.(I'm concerned about the risk, but i'm interested in the high-dividend. I wonder what the investment period and how much the minimum investment.)
87	oooo는 투자회사로서 꽤 매력적이다 그러나 채권은 기대수익율에서 주식보다 매우 낮다.(oooo is attractive as an investment company. However, bonds are much lower than stocks at expected return.)
⋮	⋮
157	위험부담이 있을 것 같아서 그런 위험을 감수하고 투자하고 싶지 않다. 원금보장이 된다면 투자의향이 있다.(I don't want to invest because it seems to be a risk. If the principal is guaranteed, i will invest.)

<Table 9>는 본 연구에서 활용한 판별알고리즘 중 Random Forest를 이용하여 예측 성능을 정오분류표로 나타낸 것이다. 각 요인 별로 실제 사용자의 텍스트데이터를 이용하여 예측하였다. 금융소비자 예측을 통한 결과는 <Table 10>과 같이 나타난다.

금융 서비스의 다양화와 고객들의 신용위험에 대한 관심이 증가하면서, 비대면 서비스가 증가하고 차별화된 서비스를 제공함으로써 고객 만족 및 서비스 만족을 증대시키고, 실시간 대응성에 따라 고객 편의 및 서비스품질을 향상시킬 수 있다.



〈Figure 3〉 Examples of Recommended Fund

〈Table 9〉 Confusion Matrix

internal attribution	experience	information	none	Predicted
Experience	1	0	0	1
Information	3	41	7	51
None	12	17	76	105
Actual	16	58	83	157

external attribution	reactivity	reliability	None	Predicted
Reactivity	8	0	2	10
Reliability	0	7	2	9
None	15	9	114	138
Actual	23	16	118	157

portfolio theory	profitability	stability	None	Predicted
Profitability	22	8	7	37
Stability	5	16	4	25
None	13	24	58	95
Actual	40	48	69	157

〈Table 10〉 Classification Result and Financial Product Recommendation Using Unstructured Data

	Text	class.In	class.Ex	class.Po	name	explanation
1	High, investment, interest, number, profit, ho, something, etc.	Information	none	stability	Long-term Securities Investment Trust 0000	The largest fund of public bond type that invests in domestic high-yield bonds and pursue long-term stable profitability
2	Think, stock, trust, fund, personal pension, general, etc.	none	reactivity	none	high-dividend focus 30 Securities investment trust 1(mix bond) S	A mixed fund that invests in high dividend stocks and invests in high-yield bonds to increase stability
⋮	⋮	⋮	⋮	⋮	⋮	⋮
86	Investment, risk, amount, minimum, per, Attention, High light, etc.	information	none	none	high-dividend focus 30 Securities investment trust 1(mix bond) S	A mixed fund that invests in high dividend stocks and invests in high-yield bonds to increase stability
87	Investment, company, stock, attractive, profit, expectation, etc.	none	reliability	profitability	Value focus 30 Securities investor trust (mix bond) S	Capital gains from high-yield bonds investments and selected value stocks funds seeking share price rises and profit from stock investment
⋮	⋮	⋮	⋮	⋮	⋮	⋮
157	Risk, investment, guarantee, intent, burden, principal, thing, same, etc	none	none	stability	Long-term Securities Investment 0000	The largest fund of public bond type that invests in domestic high-yield bonds and pursue long-term stable profitability

상품을 공급하는 판매자 입장에서도 비싼 비용을 지불하고 수요자의 정보에 관한 조사를 진행해 왔으나, 고객의 텍스트 데이터 분석을 통해, 고객의 성향분석 및 고객의 동적인 변화에 대한 신속한 대응으로 금융소비자의 수요 예측이 가능할 것이다. 특히 금융소비자가 작성한 문서를 통해 서비스수요자의 금융소비성향을 예측한다면, SNS의 Posting, 댓글, 블로그 등을 통해 금융소비자의 금융소비성향을 알 수 있다는 점에서 금융기관의 신용위험관리 및 고객관리에 효과적이고 효율적으로 수행하기 위한 지능형 기술에 대한 연구의 일환으로 볼 수 있다.

4.4 소비자 대상 추천 정확도 검증

본 연구에서 제안한 방법에 대해 추가 검증하기 위하여 추론 알고리즘과 추천 금융상품 DB로 경영 전공자들을 대상으로 실험해 보았다.

첫째, 금융소비자인 실험대상자가 투자에 대해서 남긴 글(VOC, 이전상담결과, 댓글 등 가용한 로그 데이터)을 수집하여 검증용 문장으로 삼는다.

둘째, 검증용 문장으로 형태소 분석, stopword 제거, 명사 추출 등의 전처리 작업을 수행한다.

셋째, 추출된 단어들 각각의 통계치(TF, IDF, TFIDF 등)를 계산하고 그것으로 성능분석결과 가장 우수한 것으로 선정된 판별 알고리즘을 활용하여 실험대상자의 투자성향을 추론한다.

마지막으로, 추론한 투자성향을 실험대상자에게 추천하고 그 정확성과 만족도를 점검한다.

<Table 11>은 실험대상자의 글을 기준으로 Random Forest 알고리즘을 이용하여 예측한 텍스트 데이터 자동분류 결과와 MBA 전공자들이 투자성향에 대한 분류질문을 받고 답한 것에 대한 정확도를 비교하여 정오분류표로 나타낸 것이다. F-score가 0.87 0.72 0.86으로 금융소비자 추천 알고리즘으로 평가한 것이 F-score가 0.49 0.42 0.50으로 MBA 전공자의 평가보다 본 연구에서 실시한 자동분류 성과가 뛰어났다.

〈Table 11〉 Verification of Recommendation Accuracy

RandomForest classification result table

internal attribution	Eexperience	Information	None	Predicted
Experience	4	0	0	4
Information	0	1	0	1
None	4	0	11	15
Actual	8	1	11	20

external attribution	Reactivity	Reliability	None	Predicted
Reactivity	1	0	0	1
Reliability	0	1	0	1
None	3	1	14	18
Actual	4	2	14	20

portfolio theory	Profitability	Stability	None	Predicted
Profitability	5	0	0	5
Stability	1	5	1	7
None	1	0	7	8
Actual	7	5	8	20

Investor classification result table

internal attribution	Experience	Information	None	Predicted
Experience	20	0	19	39
Information	4	5	15	24
None	24	1	32	57
Actual	48	6	66	120

external attribution	Reactivity	Reliability	None	Predicted
Reactivity	8	1	14	23
Reliability	4	6	29	39
None	12	5	41	58
Actual	24	12	84	120

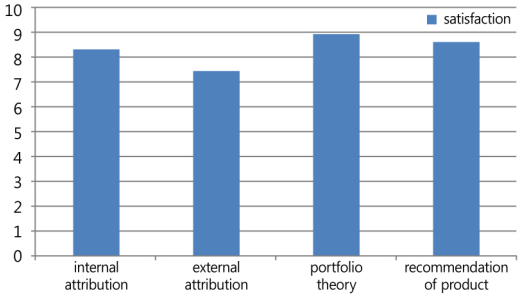
portfolio theory	Profitability	Stability	None	Predicted
Profitability	18	1	8	27
Stability	12	17	17	46
None	12	12	23	47
Actual	42	30	48	120

Comparison of classification performance

		Accuracy	Precision	Recall	F-score
Method	internal attribution	0.8000	0.9111	0.8333	0.8705
	external attribution	0.8000	0.9259	0.5833	0.7157
	portfolio theory	0.8500	0.8631	0.8631	0.8631
Person (MBA major)	internal attribution	0.4750	0.4275	0.5783	0.4916
	external attribution	0.4583	0.4029	0.4405	0.4208
	portfolio theory	0.4833	0.5085	0.4915	0.4999

4.5 소비자 대상 추천 만족도 검증

최종적으로 설문을 실시하여 MBA 전공자들을 대상으로 실험하였으며, MBA 전공자들은 실무 경험이 있고 재무 분야에 대한 배경지식이 있는 자들이다. 실험만족도는 평균을 이용하여 종합하였다. 투자와 관련된 설문의 글을 받고 그에 따른 추천만족도 평가를 시행하였다. 추천된 투자성향과 투자상품 목록을 보여준 이후에, 만족도 조사하였다. 금융소비자의 만족도는 두 가지로 투자성향, 추천상품으로 나누어서 측정했다. 우선 1차로 MBA 전공자들을 대상으로 설문자의 투자성향이 드러나는 비정형(text) 설문을 실시하였고, 2차로는 본 연구에서 제안한 텍스트분석 결과를 통하여 추천된 투자성향과 제안된 투자상품에 대한 만족도를 설문을 실시하였다. 실시한 결과 8.29, 7.43, 8.86, 8.57로 높은 만족도를 보였다.



〈Figure 4〉 Satisfaction based on Classification and Product Recommendation

5. 시사점 및 결론

본 연구에서는 Web에서 금융소비자의 비정형 텍스트 데이터만으로도 소비자의 투자성향을 자동 추론하여 금융상품서비스를 제안할 수 있는 방법론을 제안하였으며 이를 로보 어드바이저에 적용할 수 있음을 보였다.

금융소비자의 투자성향을 예측하기 위해서 Decision Tree, Logistic Regression, Support Vector Machine 등의 데이터 마이닝 분류 기법들이 사용되었다. 이러한 분류 기법들은 고객의 정적정보 즉, 인구통계학적인 정보 등을 바탕으로 기계학습을 통하여 분류 모델을 생성한다. Web을 통해 나온 금융소비자의 글을 토대로 하여, 금융소비자의 유형화를 보다 정확하게 보정하는 방안을 제시하였다. 또한 실제 금융업계에서 사용할 수 있는 예측 모델의 실시간 갱신을 가능하게 함으로써, 고객의 상태 변화에 민감하게 반응할 수 있는 기반을 제공한다.

기존 자산관리사인 PB와 달리 현재의 로보 어드바이저는 금융소비자의 투자성향에 대해 목적과 목표 같은 표준화된 질의응답을 통해 정보를 수집한다. 로보 어드바이저를 이용하는 금융소비자가 모든 정보를 직접 입력함으로써 투자위험성향도 본인의 주관적 판단에 의해 입력하게 된다. 이러한 점에서 사람에 의한 대면거래에서 이루어지는 투자위험성향과는 차이가 있게 된다. 비대면채널에서의 질문에 대해 금융소비자가 잘못 이해하거나, 제공되는 거래조건들을 제대로 이해하지 못한 경우에는 입력한 정보들이 알고리즘 하에서 어떻게 어느 정도로 고려되는지 여부는 금융소비자는 알 수가 없는데 비해 본 연구에서 제안하는 금융소비자 유형화는 이러한 부분들에 대해 고려하여 금융소비자 유형을 자동으로 분류하는 방법을 제안하였다. 분류 성능도 성공적으로 제시하였다.

추후 연구 이슈로는 첫째, 본 연구에서 다룬 금융소비자 성향분석에 대하여, one class classification을 적용하여 금융소비자의 성향에 대해 분석을 하였는데 효과적인 예측을 위해서는 분류 모형 자체의 성능을 개선하는 노력이 필요한데, diversity and accuracy-based classifier selection, data imbalance등의 주제들에 대한 이슈들이 고려되지 못하였다. 여러 추가적인 요인들이 함께 종합적으로 검토되어야 한다.

둘째, SST에 대한 영향을 미치는 요소로 금융소비자의 사회인구학적 특성이나 이용시간, 자산규모, 가계경제 특성, 서비스 이용경험 등의 요인들이 독립변수나 조절변수로 작용할 수 있기 때문에

이러한 부분들을 고려하여 지속적인 문헌 고찰과 실증 분석을 통한 확장연구를 지속해야 한다.

셋째, 일부 금융소비자를 대상으로 한정하여 연구를 진행하였기 때문에 연구대상과 표본의 대표성 문제가 제기될 수 있어 일반화 하는데 다소 한계가 있을 것이다. 따라서 향후에는 Web을 통하여 더 많은 금융소비자의 비정형데이터와 다양한 SST기반의 금융상품 데이터를 확보하여, 금융소비자 예측에 효과적으로 적용될 수 있는지 다시 한번 검증해야 할 필요가 있다.

References

Andersen, O., "On the Internationalization Process of Firms : A Critical Analysis", *Journal of International Business Studies*, Vol. 24, No.2, 1993, 209-231.

Anderson, P.F., "Marketing, Strategic Planning and the Theory of the Firm", *The Journal of Marketing*, Vol.46, No.2, 1982, 15-26.

Ban, J.I., "Investor Characteristics, Financial Literacy, and Fund Investing Behavior : Evidence from a Mutual Fund Investor Survey", *Korean Financial Management Association*, Vol.30, No.4, 2013, 1-37.

(반주일, "투자자특성, 금융이해력, 펀드투자행태에 관한 연구 : 펀드투자자 설문조사를 통한 검증", *재무관리연구*, 제30권, 제4호, 2013, 1-37.)

Barniv, R., A. Agarwal, and R. Leach, "Predicting the Outcome Following Bankruptcy Filing : a Three-State Classification Using Neural Networks", *Intelligent Systems in Accounting, Finance and Management*, Vol. 6, No.3, 1997, 177-194.

Bell, T.B., G.S. Ribar, and J. Verchio, "Neural Nets Versus Logistic Regression : a Comparison of Each Model's Ability to Predict Commercial Bank Failures", *Proceedings of the 1990 Deloitte and Touche/University of Kansas Symposium of Auditing Problems, Lawrence*, Kansas, 1990, 29-58.

Bertaut, C.C. and M. Starr, "Household Portfolios in the United States", 2000.

Betterment, "*Pricing*", 2016, Available at https://www.betterment.com/pricing (Accessed September 11, 2016).

Bharadwaj, A., O.A. El Sawy, P.A. Pavlov, and N.V. Venkatraman, "Digital Business Strategy : Toward a Next Generation of Insights", *MIS Quarterly*, Vol.37, No.2, 2013, 471-482.

Bitner, M.J., A.L. Ostrom, and M.L. Meuter, "Implementing Successful Self-Service Technologies", *The Academy of Management Executive*, Vol.16, No.4, 2002, 96-109.

Bondt, W.F. and R. Thaler, "Does the Stock Market Overreact?", *The Journal of Finance*, Vol.40, No.3, 1985, 793-805.

Breiman, L., "Random Forests", *Machine Learning*, Vol.45, No.1, 2001, 5-32.

Burges, C.J., "A Tutorial on Support Vector Machines for Pattern Recognition", *Data Mining and Knowledge Discovery*, Vol.2, No. 2, 1998, 121-167.

Campbell, J.Y., "Household Finance", *The Journal of Finance*, Vol.61, No.4, 2006, 1553-1604.

Cha, H.W., H.S. Kim, and J.Y. Bae, "The Relationship among Financial Groups' Ideal Image, Source Credibility, Perception of Issue Importance, and Financial Groups' Reputation", *Journal of Public Relations Research*, Vol.11, No.2, 2007, 89-125.

(차희원, 김훈순, 배지양, "금융권 이상적 이미지, 정보원 신뢰도 및 이슈 중요도인식과 금융기업명성 간 관련성", *홍보학연구*, 제11권, 제2호, 2007,

89-125.)

Chan, K.C., "On the Contrarian Investment Strategy", *Journal of Business*, Vol.61, No.2, 1988, 147-163.

Chang, Y., S.F. Wong, H. Lee, and S.P. Jeong, "What Motivates Chinese Consumers to Adopt FinTech Services : A Regulatory Focus Theory", *Proceedings of the 18th Annual International Conference on Electronic Commerce : e-Commerce in Smart connected World. ACM*, 2016, 40.

Choi, A.Y., Y.J. Jang, and S.M. Han, "A Study on Financial Consumer's Online Information Search Behavior According to Their Perceived Difficulties", *Korea Consumer Agency*, Vol.27, No.2, 2016, 203-231.

(최아영, 장연주, 한상만, "금융소비자의 온라인 정보탐색 어려움 유형에 따른 온라인 정보탐색행동", *소비자학연구*, 제27권, 제2호, 2016, 203-231.)

Choi, H.C. and H.J. Cho, "A Study on the Effects of Financial Education and Financial Experiences on the Financial Literacy of Undergraduate Students", *Korean Society of Consumer Studies*, Vol.22, No.4, 2011, 55-75.

(최현자, 조혜진, "금융교육과 금융거래경험이 대학생의 금융이해력에 미치는 영향에 관한 연구", *소비자학연구*, 제22권, 제4호, 2011, 55-75.)

Chordia, T. and L. Shivakumar, "Momentum, Business Cycle, and Time-Varying Expected Returns", *The Journal of Finance*, Vol.57, No.2, 2002, 985-1019.

Cortes, C. and V. Vapnik, "Support-Vector Networks", *Machine Learning*, Vol.20, No.3, 1995, 273-297.

Cox, D.R., "The Regression Analysis of Binary Sequences", *Journal of the Royal Statistical Society,* Series B, 1958, 215-242.

Dabholkar, P.A. and R.P. Bagozzi, "An Attitudinal Model of Technology-Based Self-Service : Moderating Effects of Consumer Traits and Situational Factors", *Journal of the Academy of Marketing Science*, Vol.30, No.3, 2002, 184-201.

Dholakia, R.R., M. Zhao, and N. Dholakia, "Multichannel Retailing : A Case Study of Early Experiences", *Journal of Interactive Marketing*, Vol.19, No.2, 2005, 63-74.

Ganzach, Y., "Judging Risk and Return of Financial Assets", *Organizational Behavior and Human Decision Processes*, Vol.83, No.2, 2000, 353-370.

Godin, G., P. Valois, L. Lepage, and R. Desharnais, "Predictors of Smoking Behaviour : an Application of Ajzen's Theory of Planned Behaviour", *British Journal of Addiction*, Vol. 87, No.9, 1992, 1335-1343.

Griffiths, T.L. and M. Steyvers, "Finding Scientific Topics", *Proceedings of the National Academy of Sciences*, Vol.101, No.1, 2004, 5228-5235.

Ha, Y.W. and Y.D. Kim, "The Influence of Regulatory Focus on Financial Consumers' Investment Behavior", *Korean Academic Society of Business*, Vol.40, No.3, 2011, 605-631.

(하영원, 김영두. "조절초점이 금융소비자의 투자 행동에 미치는 영향", *경영학연구*, 제40권, 제3호, 2011, 605-631.)

Hale, D. and R. Thakur, "Self-Service Technologies and Generation Y : are they Beyond the Personal Touch?", *International Journal of Electronic Marketing and Retailing*, Vol.1, No.2, 2006, 132-154.

Han, M.H., S.K. Shin, and S.A. An, "Risk Attitude and Risk-Aversion Pruchase Behavior

in the Financial Market", *Korean Academic Society of Business*, Vol.29, No.3, 2000, 337-352.

(한민희, 신상근, 안성아, "금융고객의 위험태도와 위험대응 구매행동", *경영학연구*, 제29권, 제3호, 2000, 337-352.)

Heider, F., *The Psychology of Interpersonal Relations,* New York : Wiley, 1958.

Huh, K.O. and S.H. Yoo, "Investigation of Discriminating Variables between Three Groups Divided by Risk Acceptance in Financial Investment and Finding Differences in Financial Characteristics of Household", *Financial Planning Review*, Vol.5, No.1, 2012, 133-159.

(허경옥, 유수현, "가계투자에서의 위험감수성향 판별변수조사 및 위험감수성향에 따른 가계재무관리 특성 분석", *한국FP학회*, 제5권, 제1호, 2012, 133-159.)

Huh, K.O., H.J. Lee, and M.H. Choi, "Analysis on Financial Asset Portfolio Structure in Relation to Household Head's Tendency for Financial Risk Tolerance", *Financial Planning Review*, Vol.3, No.2, 2010, 105-134.

(허경옥, 이현진, 최미향, "가구주의 위험감수성향의 변화와 금융자산 포트폴리오 구조 분석", *한국FP학회*, 제3권, 제2호, 2010, 105-134.)

Hur, J.O. and C.H. Kim, "A Study on the Marketing Strategy for Enhancing Service Quality of Local Bank", *Journal of Marketing Management Research,* Vol.6, No.3, 2001, 157-187.

(허정옥, 김정희, "연구논문: 지방은행의 서비스품질 제고를 위한 전략 모색", *마케팅관리연구*, 제6권, 제3호, 2001, 157-187.)

Hwang, J.J., "Improper Sales of Financial Products and the Direction of Policy Protecting Consumers", Master Thesis, Korea University, 2010.

(황진자, "금융상품 불완전판매와 소비자보호정책", *고려대학교 박사학위 논문*, 2010.)

Jegadeesh, N. and S. Titman, "Returns to Buying Winners and Selling Losers : Implications for Stock Market Efficiency", *The Journal of Finance*, Vol.48, No.1, 1993, 65-91.

Jiang, S., G. Pang, M. Wu, and L. Kuang, "An Improved K-nearest-neighbor Algorithm for Text Categorization", *Expert Systems with Applications*, Vol.39, No.1, 2012, 1503-1509.

Joachims, T., "Text Categorization with Support Vector Machines : Learning with many Relevant Features", *European Conference on Machine Learning*, Springer Berlin Heidelberg, 1998.

Johanson, J. and J.E. Vahlne, "The Internationalization Process of the Firm-a Model of Knowledge Development and Increasing Foreign Market Commitments", *Journal of International Business Studies*, Vol.8, No.1, 1977, 23-32.

Joo, S.H., "Workplace Financial Education and Retirement Planning", *Korean Society of Consumer Studies*, Vol.19, No.1, 2008, 71-89.

(주소현, "종업원 재무교육과 은퇴설계", *소비자학연구*, 제19권, 제1호, 2008, 71-89.)

Joo, S.H., H.C Choi, and Y.A. Sung, "Mutual Fund Investment by Individual Investors in Korea", *Financial Planning Review*, Vol.2, No.1, 2009, 75-102.

(주소현, 최현자, 성영애, "우리나라 펀드투자자의 펀드투자 의사결정에 따른 투자 행태", *한국FP학회*, 제2권, 제1호, 2009, 75-102.)

Joo. S.H., "Mutual Fund Investments Literacy of Mid-to Old-Age Consumers", *Korea Consu-*

mer Agency, Vol.40, No.1, 2011, 47-69.
(주소현, "중고령층 금융소비자의 펀드투자에 관한 이해력", *소비자문제연구*, 제40권, 제1호, 2011, 47-69.)

LearnVest, "*What you Get*", 2016, Available at https://www.learnvest.com (Accessed September 11, 2016).

Lee, H.D., S.B. Kim, and E.J. Yoo, "Relationship between Macroeconomic Variables and Fund Flow in the Korean Equity Funds", *Korean Association of Industrial Business Administration*, Vol.29, No.1, 2014, 175-197.
(이현덕, 김상배, 유은정, "국내 거시경제변수와 주식형펀드 현금흐름의 관계", *경영연구*, 제29권, 제1호, 2014, 175-197.)

Lee, S.S., "A Study on the Relationship of Service Quality, Customer Satisfaction, and Cross Buying Intension in Bank Service : Comparison of the Korea and China Bank", *Korean Corporation Management Review*, Vol. 41, No.1, 2012, 83-106.
(이상석, "연구논문 : 은행서비스에서의 서비스품질, 고객만족도 및 교차구매의도 간의 관계연구; 한, 중 은행 간 비교", *기업경영연구*, 제41권, 제1호, 2012, 83-106.)

Liljander, V., F. Gillberg, J. Gummerus, and A. Van Riel, "Technology Readiness and the Evaluation and Adoption of Self-Service Technologies", *Journal of Retailing and Consumer Services*, Vol.13, No.3, 2006, 177-191.

Lin, J.S.C. and P.L. Hsieh, "The Role of Technology Readiness in Customers' Perception and Adoption of Self-Service Technologies", *International Journal of Service Industry Management*, Vol.17, No.5, 2006, 497-517.

Lyon, L. and G.T. Cameron, "A Relational Approach Examining the Interplay of Prior Reputation and Immediate Response to a Crisis", *Journal of Public Relations Research*, Vol.16, No.3, 2004, 213-241.

Manevitz, L.M. and M. Yousef, "One-Class SVMs for Document Classification", *Journal of Machine Learning Research*, Vol.2, 2001, 139-154.

Market Riders, "*Do-It-Yourself Investing Software, Made Simple*", 2016, Available at http://www.marketriders.com (Accessed September 11, 2016).

Markowitz, H., "Portfolio Selection", *The Journal of Finance*, Vol.7, No.1, 1952, 77-91.

Meuter, M.L., A.L. Ostrom, R.I. Roundtree, and M.J. Bitner, "Self-Service Technologies : Understanding Customer Satisfaction with Technology-Based Service Encounters", *Journal of Marketing*, Vol.64, No.3, 2000, 50-64.

Meuter, M.L., A.L. Ostrom, M.J. Bitner, and R. Roundtree, "The Influence of Technology Anxiety on Consumer Use and Experiences with Self-Service Technologies", *Journal of Business Research*, Vol.56, No.11, 2003, 899-906.

Min, K.S. and I.S. Song, "Measuring the Distribution of Korean Investors Risk Aversion", *Korean Consumption Culture Association*, Vol.17, No.1 2014, 69-88.
(민경실, 송인성, "국내 금융 투자자들의 위험 회피 성향 측정", *소비문화연구*, 제17권, 제1호, 2014, 69-88.)

Mishra, S.K. and M. Kumar, "How Mutual Fund Investors' Objective and Subjective Knowledge Impacts their Information Search and Processing Behaviour", *Journal of Financial Services Marketing*, Vol.16, No.1, 2011, 27-41.

Ngai, E.W.T., Y. Hu, Y.H. Wong, Y. Chen, and

X. Sun, "The Application of Data Mining Techniques in Financial Fraud Detection : A Classification Framework and an Academic Review of Literature", *Decision Support Systems*, Vol.50, No.3, 2011, 559-569.

Personal Capital, "*What It Costs*", 2016, Available at https://www.personalcapital.com/wealth-management (Accessed September 11, 2016).

Quinlan, J.R., "C5. 0 Data Mining Tool", *RuleQuest Research*, 1997.

Quinlan, J.R., "Improved Use of Continuous Attributes in C4.5", *Journal of Artificial Intelligence Research*, 1996, 77-90.

Reinders, M.J., P.A. Dabholkar, and R.T. Frambach, "Consequences of Forcing Consumers to Use Technology-Based Self-Service", *Journal of Service Research*, Vol.11, No.2, 2008, 107-123.

Sia, S.K., C. Soh, and P. Weill, "How DBS Bank Pursued a Digital Business Strategy", *MIS Quarterly Executive*, Vol.15, No.2, 2016, 105-121.

Song, M. and S.Y. Kim, "Detecting the Knowledge Structure of Bioinformatics by Mining Full-Text Collections", *Scientometrics*, Vol.96, No.1, 2013, 183-201.

Sung, H.S. and S.L. Han, "Self-Regulatory Mode Effects on Emotion and Customer's Response in Failed Services- Focusing on the Moderate Effect of Attribution Processing", *Asia Marketing Journal*, Vol.12, No.2, 2010, 83-110.

(성형석, 한상린, "고객의 자기조절성향이서비스 실패에 따른 부정적 감정과 고객반응에 미치는 영향-귀인과정에 따른 조정적 역할을 중심으로", *아시아마케팅저널*, 제12권, 제2호, 2010, 83-110.)

Susskind, A.M., "Electronic Commerce and World Wide Web Apprehensiveness : An Examination of Consumers' Perceptions of the World Wide Web", *Journal of Computer-Mediated Communication* Vol.9, No.3, 2004.

Tan, S., "An Effective Refinement Strategy for KNN Text Classifier", *Expert Systems with Applications*, Vol.30, No.2, 2006, 290-298.

Walker, R.H., M. Craig-Lees, R. Hecker, and H. Francis, "Technology-Enabled Service Delivery : an Investigation of Reasons Affecting Customer Adoption and Rejection", *International Journal of Service Industry Management*, Vol.13, No.1, 2002, 91-106.

Warner, N. and S. Cramer, "Saving Behaviours : First Wave Baby Boomers", *Journal of Consumer Studies and Home Economics*, Vol. 19, No.1, 1995, 57-67.

WealthFront, "*Wealthfront Investment Methodology White Paper*", 2016, Available at WealthFront https://research.wealthfront.com/whitepapers/investment-methodology, (Accessed September 11, 2016).

Wu, B., T. Abbott, D. Fishman, W. McMurray, G. Mor, K. Stone, and H. Zhao, "Comparison of Statistical Methods for Classification of Ovarian Cancer Using Mass Spectrometry Data", *Bioinformatics*, Vol.19, No. 13, 2003, 1636-1643.

Zavgren, C.V., "Assessing the Vulnerability to Failure of American Industrial Firms : a Logistic Analysis", *Journal of Business Finance and Accounting*, Vol.12, No.1, 1985, 19-45.

Zeelenberg, M. and R. Pieters, "Beyond Valence in Customer Dissatisfaction : A Review and New Findings on Behavioral Responses to Regret and Disappointment in Failed Ser-

vices", *Journal of Business Research*, Vol. 57, No.4, 2004, 445-455.

Zeithaml, V.A. and M.C. Gilly, "Characteristics Affecting the Acceptance of Retailing Technologies : A Comparison of Elderly and Nonelderly Consumers", *Journal of Retailing*, Vol.63, No.1, 1987, 49-68.

Zhou, R. and M.T. Pham, "Promotion and Prevention across Mental Accounts : When Financial Products Dictate Consumers and Investment Goals", *Journal of Consumer Research*, Vol.31, No.1, 2004, 125-135.

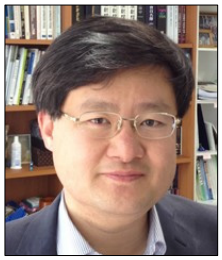
◈ About the Authors ◈

Jaewoong Lee (jw_lee@khu.ac.kr)

Jaewoong Lee has completed a master degree major in Bigdata management from Kyung Hee University. His current research interests focus on big data analysis, Economics of Information systems, Text mining, and a complex systems.

Young-Sik Kim (kys997@khu.ac.kr)

Young-Sik Kim is a researcher in the Management Research Institute Kyung Hee University. He received his master's and Ph.D. degrees in finance from Kyung Hee University. His current research interests are in the area of corporate governance, specifically boards of directors and their relationship with the CEO.



Ohbyung Kwon (obkwon@khu.ac.kr)

Ohbyung Kwon is a professor in the School of Management at Kyung Hee University. He received his master's and Ph.D. degrees in MIS from Korea Advanced Institute of Science and Technology (KAIST), and a B.A. from Seoul National University. He published several articles in Journal of MIS, International Journal of Information Management, Decision Support Systems, etc. His research interests include ubiquitous computing, big data analytics, and decision support systems.