韓國IT서비스學會誌
第17卷 第2號
2018年 6月, pp.207-218

Journal of Information Technology Services
https://doi.org/10.9716/KITS.2018.17.2.207

마케팅 관점으로 본 빅 데이터 분석 사례연구 : 은행업을 중심으로

박성수* · 이건창**

Big Data Analytics Case Study from the Marketing Perspective : Emphasis on Banking Industry

Sung Soo Park* · Kun Chang Lee**

■ Abstract ■

Recently, it becomes a big trend in the banking industry to apply a big data analytics technique to extract essential knowledge from their customer database. Such a trend is based on the capability to analyze the big data with powerful analytics software and recognize the value of big data analysis results. However, there exits still a need for more systematic theory and mechanism about how to adopt a big data analytics approach in the banking industry. Especially, there is no study proposing a practical case study in which big data analytics is successfully accomplished from the marketing perspective. Therefore, this study aims to analyze a target marketing case in the banking industry from the view of big data analytics. Target database is a big data in which about 3.5 million customers and their transaction records have been stored for 3 years. Practical implications are derived from the marketing perspective. We address detailed processes and related field test results. It proved critical for the big data analysts to consider a sense of Veracity and Value, in addition to traditional Big Data's 3V (Volume, Velocity, and Variety), so that more significant business meanings may be extracted from the big data results.

Keyword : Big Data, 5V, Big Data Analytics, Banking Industry, Target Marketing

Submitted : February 20, 2018 1st Revision : March 24, 2018 Accepted : April 26, 2018

* 성균관대학교 경영대학 박사과정
** 성균관대학교 경영대학/삼성융합의과학원(SAIHST) 교수, 교신저자

1. 서 론

은행업의 경우 2008년 금융위기 이후 저성장 속에서 과거에 비해 전반적으로 수익이 축소되고 있다. 더욱이 저금리가 장기화 되면서 수익성 개선을 위한 고비용 수익구조의 혁신이 필요한 상황이다. 또한, IT기술의 발달로 인한 인터넷, 모바일 뱅킹의 활성화로 인해 금융기관을 찾아오는 채널의 변화가 가속화 되고 있다. 이러한 이슈를 극복하고 지속적인 성과를 달성하기 위해 빅 데이터는 최근 경영환경에서 중요한 요소로 다루어지고 있다(Manyika et al., 2011, Mervis, 2012). 따라서 빅 데이터 분석 능력을 어떻게 확보하는가에 따라 은행의 향후 수익창출과 비용감소 결과가 크게 달라질 전망이다(McKinsey, 2016). 옥스퍼드 대학교와 IBM이 빅 데이터의 실제 적용에 대해 조사한 보고서에 의하면 은행 및 금융회사의 71%는 빅 데이터의 활용이 자신의 조직에 대한 경쟁력을 높이고 있다고 보고하였다(Schroeck et al., 2012).

그럼에도 불구하고 빅 데이터의 중요성에 비해서 마케팅 분야의 연구 및 실무적 활용은 명확한 방향을 제시하지 못하고 있다(Lee, 2015). 방대한 규모의 고객층을 상대하고, 모바일 및 최신 정보기술을 사용하는 은행은 은행업의 특성에 맞는 마케팅 관점의 빅 데이터 분석의 적용 메커니즘의 수립이 필요하다(Sun et al., 2014). 높은 운영 성과 및 전략적 영향에도 불구하고 빅 데이터의 잠재력을 평가하기 위한 경험적 연구가 부족하다(Wamba et al., 2015). 따라서 본 연구는 빅 데이터 실증연구의 확산차원에서 은행의 빅 데이터 분석 기반 타겟 마케팅 실행과 성과에 초점을 두어 사례를 연구하고 다음과 같은 연구질문(RQ : Research Question)을 제시한다.

RQ1 : 은행업의 타겟 마케팅에 빅 데이터 분석의 적용과 효과를 보다 세부적인 수준에서 논하고 시사점을 제시한다.

RQ2 : 은행업의 마케팅을 위한 빅 데이터 분석의 효과적인 적용을 위한 방법론을 제시한다.

본 연구의 결과는 은행의 빅 데이터 분석의 응용 및 구현에 대한 이해와 관련 전략수립에 기여할 수 있을 것이다.

2. 이론적 배경

2.1 빅 데이터의 정의

가트너(Gartner, 2012)는 빅 데이터를 조직의 통찰력과 의사결정을 향상시키기 위해 효율적이고 혁신적인 정보처리방식을 필요로 하는 대용량(Volume), 빠른 속도(Velocity) 및 다양한(Variety) 정보 자산으로 정의하며 3V로 정의하였다. Beulke(2011)는 3V에 추가적으로 데이터의 모호성과 불완전성에 대한 정확성(Veracity)과 빅 데이터를 분석해서 찾아내야 하는 숨겨진 가치(Value)까지 포함된 5V로 정의하였다. 5V의 자세한 정의와 내용은 아래와 같다.

2.1.1 대용량

일반적으로 방대한 양의 저장소를 사용하거나 많은 수의 레코드로 구성되는 대량의 데이터를 의미한다(Davenport et al., 2012; Russom, 2011). 정보가 생성되는 원천이 계속 증가하면서 데이터 용량은 계속해서 기하급수적으로 증가하게 된다. 은행은 자금 입/출금 및 이체, 예금, 소비자 대출, 상업 대출, 모기지 론, 자동화기기 거래(ATM), 인터넷 뱅킹 및 모바일 뱅킹 로그 등과 같은 상품과 시스템에서 다양하고 방대한 데이터가 생성된다. 이러한 데이터를 저장하고 관리하기 위해 은행은 주로 관계형 데이터베이스 관리시스템(DBMS)을 사용하고 있다. 그러나 빅 데이터의 처리와 분석을 위해서는 하둡(Hadoop) 및 맵리듀스(Map Reduce)와 같은 분산/병렬 처리 기반의 분석 플랫폼이 필요하다.

2.1.2 빠른 속도

데이터 발생 빈도 및 데이터의 전달 빈도를 의미한다(Russom, 2011). IT기술의 발달과 인터넷 뱅킹이나 모바일 뱅킹과 같은 비대면 채널의 사용 증가로 디지털화된 데이터의 수집이 빠른 속도로

이루어지고 있다. 이러한 고객의 비대면 은행 업무처리의 이면에는 은행 사용자의 흔적과 행동이 데이터로 남게 된다. 은행은 시간 경과에 따라 생성, 처리 되는 데이터를 빠르게 분석 및 해석해서 실제 업무에 적용하는 속도가 중요하다.

2.1.3 다양성

다양한 데이터의 원천에서 각기 다른 형식으로 생성되고, 다차원 데이터 필드가 포함 된 데이터를 의미한다(Davenport et al., 2012; Russom, 2011). 거래 기록과 같이 은행의 내부에서 생성되는 데이터 외에도 은행의 외부에서 발생되는 소셜 미디어(페이스 북, 트위터, 블로그, 포럼 등), 지형 공간 데이터 및 기타 공개된 공공정보를 포함한다. 이러한 데이터들은 크게 3가지 -(i) 데이터베이스 내의 행과 열로 표현 가능한 구조화 된 데이터 (ii) 완전하게 규칙적이지는 않지만 관계형 구조 또는 특성을 가진 반 구조화 된 데이터 (iii) 분석 또는 조회를 위해 관계형 테이블로 쉽게 색인 될 수 없는 구조화되지 않은 데이터- 로 구분할 수 있다.

2.1.4 정확성

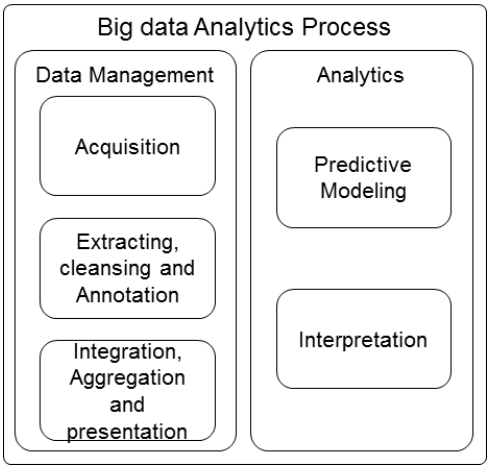
데이터의 신뢰성으로 표현하기도 하며, 오류 데이터의 선별 및 관련성 있는 데이터의 생성을 의미한다(Beulke, 2011). 빅 데이터의 또 다른 중요한 속성은 특정 유형의 데이터와 관련된 불확실성이 낮아야 한다는 것이다. 높은 데이터 품질은 빅 데이터 분석을 적용하기 위해 중요한 요구 사항이다(Schroeck et al., 2012). 따라서 검증되고 관련성 있는 데이터를 수집하고 불량 데이터를 차단하는 역량이 필요하다(Beulke, 2011).

2.1.5 가치

빅 데이터 분석이 지식, 혁신 및 향상된 의사결정 프로세스의 발전을 위한 디딤돌로써 조직에 미치는 영향의 중요성을 인식해야 한다(McAfee et al., 2012). 빅 데이터의 전략 수립 시 정보 혜택의 관점까지 고려해서 비즈니스 자산, 인적 자원 외에도 빅 데이터 자산을 더 잘 활용하여 가치를 창출하는 것이 기업의 경쟁 우위를 확보할 수 있는 기반이 된다(Wamba et al., 2015; Wixom et al., 2013).

2.2 빅 데이터 분석과 절차

빅 데이터 분석(Big Data Analytics)은 높은 볼륨, 속도 및 다양성의 관점에서 빅 데이터를 처리하기 위한 기술을 포괄하는 용어(Duan and Xiong, 2015; Wang et al., 2016; Zhou et al., 2014)로 한정지어 정의하는 연구자도 있다. 하지만, 기업에서 빅 데이터의 잠재 가치는 비즈니스 의사결정을 유도하기 위해 활용 될 때 진가가 드러나게 된다. 5가지 빅 데이터의 정의 -대용량, 빠른 속도, 다양성, 정확성, 가치- 에 따라 빅 데이터 분석을 다른 말로 표현하면, “기존 데이터 관리 기술의 한계로 인해 활용되지 않은 데이터의 저장, 처리 및 분석으로 더 높은 수준의 의사결정을 실현하는 것”으로 정의할 수 있다. 또한, 빅 데이터 분석 기반의 의사결정을 위해서는 많은 양의 다양한 데이터를 빠르게 처리 및 분석하고 의미 있는 지식으로 변환하는 효율적인 절차가 필요하다. Labrinidis and Jagandish (2012)는 빅 데이터에서 통찰력을 지닌 가치를 추출하는 전반적인 과정을 <Figure 1>과 같이 제시



〈Figure 1〉 Processes for Extracting Insight from Big Data

하였다. 빅 데이터 분석 절차는 크게 데이터 관리(Data management)와 분석(Analytics)의 두 가지로 구성된다. 데이터 관리는 1)데이터의 수집, 2)추출 및 정제, 3)통합 및 요약의 3단계로 나눌 수 있고, 분석은 빅 데이터 기반의 4)예측 모형과 그 결과에 대한 5)해석으로 나누어진다.

2.2 빅 데이터 분석을 통한 마케팅 개선

빅 데이터 분석은 마케팅 부분에서 교차 판매, 고객 마이크로 세분화, 감정 분석, 다 채널 소비자 경험 강화 등에 주로 활용된다(Duan and Xiong, 2015; Wang et al., 2016; Zhou et al., 2014). 빅 데이터 분석을 마케팅에 적용하기 위한 방향으로 Schroeck et al.(2012)는 마케팅에서 빅 데이터의 활용을 위해 다섯 가지 필요 요소를 제시하였다. 첫째, 최초 노력은 고객 중심 산출물에 초점을 두어야 한다. 둘째, 기업의 수준에 맞는 적절한 빅 데이터 계획을 수립한다. 셋째, 빅 데이터를 활용하여 즉시적인 성과를 달성하도록 적은 규모의 기존 데이터로 시작한다. 넷째, 사업 우선순위에 따라 빅 데이터 분석 역량을 확대 적용한다. 마지막으로, 빅 데이터 분석의 적용은 성과 측정 이 가능한 사업으로 기획한다.

마케팅 적용을 위한 빅 데이터 분석은 주로 고객 행동(Customer behavior) 분석에 기반을 둔 고객의 요구(Needs)에 대응하는 연구가 주로 수행 되었다(Lee, 2015). Choi et al.(2012)은 신용카드 이용 데이터에서 소비행동 패턴으로 고객 요구의 유형화를 연구 하였다. 이를 위해 주성분분석, 연관성분석, 다차원척도법 등의 통계분석방법론을 적용하였다. 또한, 고객 요구 유형화 변수를 고객의 생애 주기(Life Stage)별 대표 요구 유형 및 다중 요구 유형으로 활용하는 방안을 제시하였다. Kim(2014)은 하둡(Hadoop)기반의 빅 데이터 분석 시스템을 사용하여 증권사의 금융 상품 추천에 대한 연구를 수행하였다. 고객 400만 명의 고객 기본 데이터와 3년 간의 상품 보유 데이터 및 채널 접촉 데이터 등을 이용하여 총 거래일, 마케팅 휴면일, 이탈 소요 기간, 최고자산 이탈 소요일, 최고자산 소요기간, Portfolio 비율 등의 고객의 행동과 상태를 나타내는 변수를 산출하였다. 이를 사용하여 고객 세분화(Customer segmentation) 및 고객군별 프로파일(Profile)을 바탕으로 각 고객에게 적시적인 금융 상품의 추천 방안을 제시하였다. Danesi and Rea(2016)은 은행에서 금융상품에 대한 선호 고객을 선별하는 것을 목적으로 특정 금융 상품에 대한 고객의 구매 경향을 모델링 하였다. 빅 데이터 관점에서 월간 고객 집계 데이터와 일일 고객 잔액 데이터와 고객의 금융 거래 행동인 거래 로그(Transaction log)를 사용하여 1,100개의 설명변수를 생성하였다. 예측 모형의 생성은 6개월 간의 데이터로 학습하고 이후 1개월 데이터로 모형의 효과를 검증하였다. 로지스틱회귀분석, 의사결정나무, 랜덤포레스트로 예측 모형을 만들어 기존 전문가 시스템 기반의 상품추천 방식에 비해 높은 성능을 보였고, 특히 ROC 기준으로 랜덤포레스트가 가장 좋은 성능을 보였다. 또한, Chang and Kim(2018)은 금융기관의 고객 세분화를 목적으로 인구통계정보, 요구불 거래내역, Web/App Log, 카드데이터 등의 금융기관이 보유한 다양한 빅 데이터를 활용하였다.

금융기관의 빅 데이터 분석의 마케팅 적용에 관한 기존 사례 연구는 빅 데이터를 마케팅에 적용하기 위한 다양한 고려사항과 최종적인 새로운 비즈니스 가치 창출을 위한 빅 데이터 활용 방안을 제시하고 있다. 하지만, 두 가지 측면에서 이전의 연구는 한계를 보인다. 1) 금융기관이 보유한 다양한 데이터 중 일부를 빅 데이터 분석에 사용하고, 2) 빅 데이터 분석의 마케팅 적용과 효과가 추상적(Xu et al., 2016)이다. 이는 빅 데이터가 현재 준비 및 도입 단계로 관련 사례나 실적 데이터가 부족한 것에 기인한다. 본 사례연구는 이전의 연구에서 부분적으로 활용된 데이터들을 통합적으로 빅 데이터를 구성하고, 이를 분석하고 활용한 타겟 마케팅 실행과 성과를 제시한다. 이를 통해 은행

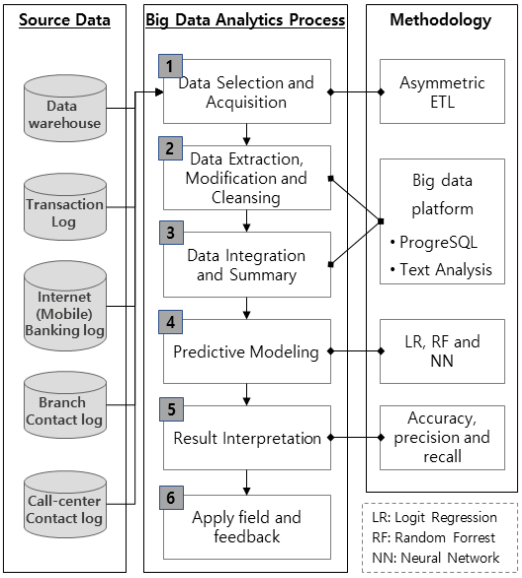
에서의 빅 데이터 분석의 마케팅 적용에 대한 시사점과 마케팅 관점의 빅 데이터 분석 방법론에 대해 논하고자 한다.

3. 사례 연구

본 연구의 대상 기업은 대한민국에서 영업 중인 예수금 기준 상위 10위 이내인 A금융기관(이하 A사)에서 수행한 빅 데이터 기반의 타겟 마케팅 개념증명(Prove of concept) 과제에 대한 사례 연구이다. A사의 기존 타겟 마케팅 방식은 고객의 연령대, 직업군, 멤버십 등의 전통적인 변수를 사용하는 규칙 기반(Rule base)의 전문가 시스템으로 구현되어 있었다. 이러한 방식은 단기적인 캠페인 실행 시 마케터의 직관에 따라 마케팅 대상 고객을 선정하는 것을 목적으로 한다. 이전 방식은 간편하면서도 빠른 목표 고객 선정이 가능하지만, 고객 요구에 맞춘 서비스를 제공할 수 없는 한계로 인하여 캠페인의 결과는 일정 수준을 넘어서지 못하는 한계를 보였다. A사는 빅 데이터 분석 기반 타겟 마케팅을 구현하여 고객의 행동에 기반한 금융 요구와 상황을 반영한 타겟 마케팅 대상 고객을 선별하고자 하였다. 본 사례연구의 빅 데이터 분석은 A사 고객 350만 명에 대한 과거 3년간의 데이터를 사용하여, 빅 데이터에서 생성한 설명변수로 종속변수를 예측하는 모형을 생성하는 것이다. 빅 데이터 분석 환경은 포스트그레씨퀄(Postgre-SQL)과 맵리듀스(Map/Reduce) 기반의 상용 빅 데이터 분석 플랫폼으로 빅 데이터의 저장, 가공, 처리, 분석을 하였으며, 텍스트 분석은 파이썬(Python)기반의 상용 텍스트 분석 도구로 분석 및 분류를 수행하였다. 빅 데이터 분석으로 고객의 금융상품 가입여력, 금융기관 및 상품에 대한 선호도, 금융기관과 고객 간의 관계를 파악한다. 최종적으로 금융상품의 교차판매(Cross-sell)가 가능한 타겟 고객을 선정하고 실제 적용을 실증한다.

3.1 빅 데이터 분석기반 타켓 마케팅 절차

본 사례 연구의 빅 데이터 분석 절차는 선행연구인 Labrinidis and Jagadish(2012)이 제시한 빅 데이터 분석 절차 5단계에 마케팅 적용을 위해 필수적인 단계를 추가하였다. 데이터 수집 단계에 데이터 선별을 포함하고, 빅 데이터 분석의 마케팅적 효과를 측정하기 위한 적용 및 피드백 단계를 추가하였다. 본 연구의 빅 데이터 분석 절차를 정리하면 <Figure 2>와 같다.



〈Figure 2〉 Block Diagram of Big Data Analytics for Target Marketing in Banking

- 1단계 : 데이터 선별과 수집-분석 목적에 적합한 데이터의 선별. 이를 수집하기 위한 기술 방안 수립 및 구현.
- 2단계 : 데이터 추출, 가공, 정제-필요 데이터의 추출. 새로운 분석 변수의 생성. 텍스트 값을 특정 범주로 분류.
- 3단계 : 데이터 통합, 요약-각 데이터를 키 변수를 기준으로 통합 및 통계 정보 생성.
- 4단계 : 예측 모형 생성-빅 데이터에 대해 높은 성능의 예측 모형 생성 및 선택.

- 5단계 : 결과의 해석 및 적용-예측 모형의 생성 결과에 대한 타당성 검토.
- 6단계 : 적용 및 피드백-분석 결과의 현장 적용. 예측 모형의 수정 및 개선.

3.2 데이터 선별과 수집

프로젝트 초기에 데이터 사이언티스트와 마케팅 담당자 간에 협의를 통해 수집할 데이터를 선별하였다. 소셜 데이터와 같은 비정형 외부 데이터는 얻을 수 있는 효과에 비해 데이터 수집 및 처리에 시간과 비용 소모가 많다고 판단하고 수집 데이터에서 제외하였다. Schroeck et al.(2012)의 제안대로 은행 내부에 신속하게 빅 데이터 분석을 적용하기 위해 대상 데이터의 범위를 5가지 데이터 원천으로 한정하고 데이터를 수집하였다. 모든 데이터는 주소, 전화번호, 이메일 주소 등의 고객식별 정보를 제거하고, 고객식별이 불가능한 통계정보만을 분석에 사용하였다.

데이터웨어하우스에서 고객 기본 정보와 고객의 월 단위 계좌 이력 데이터를 빅 데이터 플랫폼으로 수집하였다. 이후, 자금이체 로그, 인터넷(모바일) 뱅킹 로그, 대면 창구 접촉 이력, 콜센터 상담 기록을 수집하여 빅 데이터 플랫폼으로 이행하였다.

3.3 데이터 추출, 가공, 정제

데이터 추출, 가공, 정제 단계는 이전의 데이터 분석과 빅 데이터 분석이 대표적으로 차별화가 되는 부분이다. 본 연구가 이전의 상품추천과 다른 점은 Choi et al.(2012)와 Danesi and Rea (2016)가 연구에 사용한 금융거래 데이터와 Kim (2014)가 제시한 고객 접촉이력에서 추출한 고객 행동과 상태 정보를 분석 변수로 사용하였다는 점이다. 빅 데이터의 다양성 관점에서 이전에는 활용이 되지 않은 다양한 데이터에서 정보를 추출하고 분석 변수를 생성하였다. 비정형 데이터인 자금이체 로그의 통장기록 텍스트에 텍스트 분석을 적용하여 자금이체 목적을 파악하였다. 예를 들어 고객이 사용하는 신용카드 회사와 결제금액, 가입한 보험과 납부 보험료, 연금수령 등의 고객 행동과 상태를 이해할 수 있는 정보를 추출 하였다. 콜센터 상담기록 역시 텍스트 분석을 사용하여 고객의 문의, 불만 및 관심사항에 대한 정보를 추출하였다. 반정형 데이터인 인터넷(모바일) 뱅킹 로그를 파싱해서 온라인상의 고객이 방문한 페이지 정보를 추출해서 인터넷(모바일) 뱅킹 방문주기, 이용한 금융서비스, 조회한 금융상품 등의 고객의 행동 패턴을 추출하였다. 데이터 추출 단계의 주요 내용은 <Table 1>과 같다.

정형 데이터와 비정형 데이터로부터 생성된 정보를 사용하여 7가지 영역 154개의 설명변수를 생성하였다. 금융상품 가입여력 변수 55개, 은행 관여도 변수 19개와 거래패턴변수 6개로 총 80개의 고객관련 정보를 생성하고 변수로 변환하였다. 상품관련 변수는 총 74개 -상품관여도 39개, 안전자산 선호도 2개, 금융기관 접촉활동 30개, 불만사항 3개- 의 변수를 생성하였다. 본 사례의 설명변수의 의미와 주요 변수는 <Table 2>와 같다.

〈Table 1〉 Extracted Information from Major Data Source

Data source	Methods	Extracted information	
		Type of customer behavior	Pattern of customer behavior
Transaction log	Text analysis	Purpose of Fund Transfer	Cycles, periods, frequency
Internet (Mobile) banking log	Web log analysis	Type of visit page	
Call-center contact log	Text analysis	Purpose of contact	

〈Table 2〉 Description of input Variables

Variable Groups	Meaning	Major Variables
Affordable to make account	Economic allowance to purchase additional items from the bank	Closeness of financial instruments expiration date
		Increase amount of insurance fee
		Decrease amount of insurance fee
		Increase of credit card amount
		Decrease of credit card amount
Degree of engagement for bank	Closeness between customer and financial institution	Visit or contact period
		Count of contact
Transaction pattern	Specific patterns or trends in customer behavior during banking transactions	Count of transaction by purpose
		Day count after product expiration
Degree of engagement for product	Closeness between customer and financial instruments	Increase of contract period of financial instruments for saving money
		Decrease of contract period of financial instruments for saving money
		Count of fund transfer transactions
		Count of financial instruments page view
Preference for safety asset	Safety product priority purchase	Make account when get surplus funds
Contact activity	Contact Behavior by channel (ATM, internet or mobile bank, off-line branch)	Count of contact channels
		Type of task by channels
Voice of Customer	Inquiries or complaints received by the call center	Count of call center contact
		Contact type
		Type of customer complaint

3.4 데이터 통합, 요약

각 데이터 원천별로 추출 및 생성된 종속변수와 설명변수들은 예측 모형에서 학습시키기 위한 형태로 통합되었다. 본 사례의 분석 목표는 금융상품 판매 타겟 고객을 선정하는 것으로, 고객을 중심으로 변수들을 요약하고 통합하였다. 텍스트 분석으로 도출한 월 카드 결제금액, 보험료 납입액과 같은 시계열 데이터들은 시간 흐름에 따른 상승, 유지, 하락으로 단순화 하여 요약하였다. 설명 변수들 중에서 특정 값이나 구간에 95% 이상 편향되어 있는 변수, 결측치가 50% 이상인 변수와 변수간의 상관성이 0.6 이상인 변수를 제거하였다. 최종적으로 총 49개의 변수(가입여력 9개, 은행관여도 8개, 상품관여도 19개, 접촉정보 11개, 거래패턴 1개)가 예측모형의 입력으로 사용되었다. 예측모형 생성 및 검증을 위해, 3년 치 데이터 중에서 2년 6개월의 데이터를 학습데이터로 하고, 남은 6개월 데이터를 검증데이터로 사용하였다.

3.5 예측 모형 생성

본 연구에서 예측모형은 상품가입의 확률을 계산하는 것을 목적으로 수신 상품가입 여부를 종속변수로 하였다. 학습모형은 통계적인 방법인 로지스틱회귀분석(LR; Logistic Regression)과 기계학습 방법인 랜덤포레스트(RF; Random Forrest), 인공신경망(NN; Neural Network)을 사용하여 학습하였다. 각 예측 모형의 정확도는 <Table 3>과 같다. 인공신경망이 Accuracy와 Precision이 0.753과 0.75로 가장 높게 나타났다. 하지만, 본 사례연구의 목적은 금융상품 가입 고객을 선정하는 것으로, 로지스틱 회귀분석 모형이 3가지 예측모형 중 실제 상품 가입이 발생한 고객을 맞추는 비율인 Recall값이 가장 높은(0.698) 모형으로 선택되었다.

〈Table 3〉 Accuracies of Predictive Modeling for Targer Customer Selection

Modeling Methods	Accuracy	Precision	Recall
LR	0.717	0.610	0.698
RF	0.690	0.637	0.663
NN	0.753	0.750	0.645

*LR : Logistic Regression, RF : Random Forrest, NN : Neural Network.

3.6 분석 결과의 해석

생성된 타겟 마케팅 모형의 결과에 대해 데이터 사이언티스트와 은행 지점의 마케팅 및 상품 판매 담당자가 함께 해석하고 효용성을 판단하였다. 모형의 Lift 결과는 타겟 고객 지수 상위 20%의 가입률이 평균대비 1.7배 이상이었다. 로지스틱 회귀분석에서 수신 상품 가입에 영향이 없는 것으로 나타나는 변수 5개(보험 계약금액, 보험 계좌 수, 인터넷뱅킹 거래금액, 자동화기기 거래금액, 텔레뱅킹 거래건수)는 제외되었다. 빅 데이터 분석기반의 최종적인 타겟 마케팅 모형은 설명 변수는 44개로 나타났다. 이중 빅 데이터에서 생성된 설명변수 16개 변수는 <Table 4>로 정리하였다.

빅 데이터 분석 결과 몇몇 변수는 그 동안 인지하지 못한 고객 패턴을 알아낼 수 있었다. 예를 들어 퇴직금의 수령과 연금소득 발생은 수신 상품의 가입에 정(+)의 영향을 나타냈다. 이것은 은퇴 고객의 경우 상대적으로 안전성이 높은 금융상품의 가입의사가 높은 것을 나타낸다. 또한, 금융기관과 거래방식에서 창구나 자동화기기 사용이 높거나 직원 접촉이 많은 고객이 인터넷 뱅킹과 스마트 뱅킹 등의 비대면 금융 거래 위주의 고객에 비해 상품 가입에 정(+)의 영향을 미쳤다. 이것은 현재 금융기관의 패러다임이 비대면의 전자금융으로 변화해 가는 과도기에서 전통적인 금융업을 중심으로 영업하던 연구 대상 기업의 특징으로 볼 수 있다. 또한, 납입보험료가 이전 6개월 대비 감소하는 경우 수신 상품 가입에 정(+)의 영향을 미치지만 납입보험료 유지나 증가하는 경우 음(-)의 영향을 보였다.

〈Table 4〉 Input Variables from Big Data

Categories	Variables
Affordable to make account	• Increase amount of insurance fee • Decrease amount of insurance fee • Increase amount of income
Degree of engagement for bank	• Family customer • Elapsed period after contract • Elapsed period after last contact
Transaction pattern	• Retirement payment is received • Pension receipt • Public pension payment
Degree of engagement for product	• Increase of Savings product Contract Period • Decrease of Savings product Contract Period
Preference for safety asset	• N/A
Contact activity	• Number of branch visits • Count of ATM usage • Count of internet banking usage • Count of smart banking usage
Voice of Customer	• Ask product registration

3.7 분석 결과 현장 적용 및 피드백

빅 데이터 분석의 마케팅 적용 효과를 파악하기 위하여 A사의 실제 지점을 대상으로 하는 파일럿 캠페인을 수행하였다. 최근 3개월간의 데이터를 새롭게 수집하고, 로지스틱회귀분석 기반의 타겟 마케팅 모형에 입력하여 산출된 확률을 타겟 고객 지수로 계산하였다. 타겟 고객 지수 상위 20%고객 중에서 개인정보활용 및 마케팅 활용에 동의한 고객을 추출하여 지점으로 배포하였다. 수신 상품 교차판매 된 고객을 캠페인 성공 기준으로 정하고 6주간 캠페인을 진행하였다. 캠페인 결과 수신 상품 교차판매 성공률은 이전의 유사한 우수고객 대상 교차판매 캠페인의 성공률 8~10%에 비해 약 2배 이상의 성과를 달성하였다. 캠페인 결과를 사용하여 분석 모형의 조정 및 고도화를 실시하였다.

4. 연구 결과

본 연구의 목적은 은행에서 빅 데이터 분석 기반의 타겟 마케팅 사례를 통해서, 1) 은행의 타겟 마케팅에 빅 데이터 분석의 적용과 효과를 보다 세부적인 수준에서 논하고 이슈 및 시사점을 제시하는 것과 2) 은행업에서 빅 데이터 분석의 효과적인 적용을 위한 방법론을 제시는 것이다. 이를 통하여 금융기관의 마케팅 관점에서 빅 데이터 분석의 응용 및 구현에 대한 이론과 실무에 기여하고자 한다.

본 연구는 빅 데이터의 활용에 대한 실제적인 내용을 보여줌으로서 빅 데이터 분석의 수행과 적용에 좋은 본 보기가 될 것이다. 첫째, 이전의 연구들은 금융기관이 다양한 데이터 중 일부의 데이터를 사용하여 부분적인 효과를 제시하였지만 본 사례연구는 금융기관이 보유하고 있는 자금이체 로그, 인터넷/모바일 뱅킹 로그, 콜센터 상담 내역, 고객 접촉이력 등의 다양한 빅 데이터를 통합적으로 분석하였다. 둘째, 빅 데이터 분석의 결과를 실제 마케팅 활동에 적용하여 이전의 마케팅에 비해 2~3배의 높은 성과가 나타나는 것을 확인하였다.

본 연구의 사례가 주는 실무적 시사점은 다음과 같다. 빅 데이터 분석 적용의 초기 단계에는 이전에 분석에 활용되지 않던 데이터를 적용하고, 효과를 확인하며 단계적으로 빅 데이터의 대상과 범위를 확대하는 것이 바람직하다. 즉, 빅 데이터 분석을 적용하기 위해서는 기업 외부의 데이터 보다는 내부에 축적된 방대한 데이터를 우선적으로 활용하고 단계적으로 기업 외부의 정형데이터와 비정형 데이터로 데이터의 범위를 확대해 나가는 것이 효과적이다. 또한, 수집된 빅 데이터에서 최대한 많은 정보를 추출하기 위해 비정형적인 텍스트와 반정형적인 로그 데이터의 분석이 반드시 필요하다. 이러한 정보는 이전에 파악하지 못한 고객의 행동 및 상태를 이해하는 추가적인 설명변수로 활용 할 수 있고, 이를 통해 더욱 정교하고 견고한 예측 모형의 생성이 가능하다.

본 사례연구의 이론적 시사점은 다음과 같다. 첫째, 여러 연구들이 제안하고 있는 빅 데이터의 특성이 혼재되어 있는 상태에서 Beulke(2011)가 제안한 5가지 빅 데이터의 특성을 고려하는 것이 중요한 점을 밝혔다. 일반적으로 빅 데이터를 정의하는 3V(대용량, 빠른 속도, 다양성)에 정확성, 가치가 포함되어야 비즈니스 적으로 의미가 있는 빅 데이터가 된다. 본 연구에서 다룬 사례는 빅 데이터의 특성 5가지 -대용량, 빠른 속도, 다양성, 정확성, 가치- 를 충분하게 반영하였다. 대용량과 빠른 속도 면에서 이전에는 시스템의 한계로 분석이 불가능 하던 약 350만 고객의 3년간의 데이터를 분석했다. 5가지 데이터 원천에서 수집된 다양한 정형/비정형 데이터를 분석에 사용하여 다양성을 충족하였다. 사실을 나타내기에 불충분하거나 오류를 발생할 수 있는 소셜 데이터는 제외하여 정확성을 확보하고, 마지막으로 빅 데이터 분석의 결과를 실제 마케팅 활동에 적용하여 가치로 연결되는 것을 확인하였다. 둘째, 선행연구인 Labrinidis and Jagadish(2012)의 빅 데이터 분석 절차 5단계를 마케팅 적용 관점으로 수정하여 체계적인 빅 데이터 분석 방법을 제시하였다. 마케팅 활용을 위한 빅 데이터 분석 방법은 1)데이터의 선별 및 수집, 2)추출 및 가공, 3)요약 및 통합, 4)예측 모형 생성, 5)해석, 6)적용 및 피드백의 6단계가 효과적이다. 본 사례연구에 포함된 데이터의 선별과 빅 데이터 분석의 적용 및 피드백 절차는 빅 데이터 분석의 효율적 적용과 성과 측정을 위해서는 반드시 필요하다.

향후 연구는 이번 연구의 결과로 타겟팅 된 고객의 니즈에 맞는 추천 상품을 제시하기 위한 연구가 필요하다. 빅 데이터 분석 기반의 고객 세분화 및 고객군별 추천 상품의 도출에 대한 연구가 되어야 한다. 또한, 추가적인 연구는 비식별화 데이터 사용해서 타 금융기관, 통신, 교통, 유통, 공공 데이터를 융합하여 빅 데이터를 구성하고 분석하는 것이다. 이를 통하여 보다 확장된 고객 통찰력을 얻고 효율적인 마케팅 및 상품 판매 기회를

모색할 수 있을 것으로 생각된다. 또한, 빅 데이터 활용의 선결요건인 개인정보보호의 문제 및 중요성에 대한 추후 연구역시 매우 긴요할 것이다.

References

Beulke, D., "Big Data Impacts Data Management : The 5 Vs of big data", 2011, Available at http://davebeulke.com/big-data-impacts-data-management-the-five-vs-of-big-data (Accessed Jan 30. 2017).

Chang, M.S. and H.J. Kim, "A Customer Segmentation Scheme Base on Big Data in a Bank", *Journal of Digital Contents Society*, Vol.19, No.1, 2018, 85-91.

(장민석, 김형중, "빅데이터를 활용한 은행권 고객 세분화 기법 연구", *한국디지털콘텐츠학회 논문지*, 제19권, 제1호, 2018, 85-91.)

Choi, B.J., H.J. Kim, J.H. Kim, and S.H. Jin, "Data Analytics for CRM in the Age of Big Data", *Entrue Journal of Information Technology*, Vol.11, No.1, 2012, 19-27.

(최병정, 김혜진, 김자호, 진서훈, "빅 데이터 시대의 CRM을 위한 데이터 분석", *Entrue Journal of Information Technology*, 제11권, 제1호, 2012, 19-27.)

Danesi, I.L. and C. Rea, "A Customer Relationship Management Case Study Based on Banking Data", *In International Workshop on Machine Learning, Optimization and Big Data*, Springer, Cham, 2016, 224-235.

Davenport, T.H., P. Barth, and R. Bean, "How 'Big Data' is Different", *MIT Sloan Management Review*, Vol.54, No.1, 2012, 43-46.

Duan, L. and Y. Xiong, "Big data analytics and business analytics", *Journal of Management Analytics*, Vol.2, No.1, 2015, 1-21.

Gartner, "Big Data", 2012, Available at http://www.gartner.com/it-glossary/big-data/ (Accessed Jan 30. 2017).

Kim, S.S., "A study on development method for practical use of Big Data related to recommendation to financial item", *Journal of the Korea Society of Computer and Information*, Vol.19, No.8, 2014, 73-81.

(김석수, "금융 상품 추천에 관련된 빅 데이터 활용을 위한 개발 방법", *한국컴퓨터정보학회논문지*, 제19권, 제8호, 2014, 73-81.)

Labrinidis, A. and H.V. Jagadish, "Challenges and opportunities with big data", *Proceedings of the VLDB Endowment*, Vol.5, No. 12, 2012, 2032-2033.

Lee, S.K., "A Review of Big Data Analysis Based on Marketing Perspective", *Korean Journal of Business Administration*, Vol.28, No.1, 21-35, 2015.

(이서구, "빅 데이터 분석에 관한 마케팅적 접근", *대한경영학회지*, 제28권, 제1호, 2015, 21-35.)

Manyika, J., M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. Byers, "Big data : The next frontier for innovation, competition, and productivity", 2011, Available at http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation (Accessed Jan 30. 2017).

McAfee, A., E. Brynjolfsson, T.H. Davenport, D. Patil, and D. Barton, "Big data. The Management Revolution", *Harvard Business Review*, Vol.90, No.10, 2012, 61-67.

McKinsey, "The future of bank risk management", 2016, Available at http://www.mckinsey.com/business-functions/risk/our-insights/the-future-of-bank-risk-management (Accessed Jan 30. 2017).

Mervis, J., "Agencies rally to tackle big data",

Science, Vol.336, No.6077, 2012, 22-22.

Russom, P., "The three Vs of big data analytics", *TDWI Best Practices Report*, Fourth Quarter, Vol.18, 2011, 1-35.

Schroeck, M., R. Shockley, J. Smart, D. Romero-Morales, and P. Tufano, "Analytics : The real-world use of big data", *IBM Global Business Services*, 2012, 1-20.

Sun, N., J.G. Morris, J. Xu, X. Zhu, and M. Xie, "iCARE : A framework for big data-based banking customer analytics", *IBM Journal of Research and Development*, Vol.58, No. 5/6, 2014, 4-1.

Wamba, S.F., S. Akter, A. Edwards, G. Chopin, and D. Gnanzou, "How 'big data'can make big impact : Findings from a systematic review and a longitudinal case study", *International Journal of Production Economics*, Vol.165, 2015, 234-246.

Wang, G., A. Gunasekaran, E.W. Ngai, and T. Papadopoulos, "Big data analytics in logistics and supply chain management : Certain investigations for research and applications", *International Journal of Production Economics*, Vol.176, 2016, 98-110.

Wixom, B.H., B. Yen, and M. Relich, "Maximizing value from business analytics", *MIS Quarterly Executive*, Vol.12, No.2, 2013, 111-123.

Xu, Z., G.L. Frankwick, and E. Ramirez, "Effects of big data analytics and traditional marketing analytics on new product success : A knowledge fusion perspective", *Journal of Business Research*, Vol.69, No.5, 2016, 1562-1566.

Zhou, Z.H., N.V. Chawla, Y. Jin, and G.J. Williams, "Big data opportunities and challenges : Discussions from data analytics perspectives", *IEEE Computational Intelligence Magazine*, Vol.9, No.4, 2014, 62-74.

About the Authors

Sung Soo Park (volf74@empas.com)

Sung Soo Park is currently a Ph.D. student in the Business School at Sungkyunkwan University, Republic of Korea. He is interested in business big data-mining, artificial intelligence based on machine learning and Human-Robot Interaction (HRI).

Kun Chang Lee (kunchanglee@gmail.com)

Kun Chang Lee is a full professor of MIS in SKK Business School at Sungkyunkwan University (Seoul, Republic of Korea). He also holds a faculty position at Department of Health Sciences and Technology in the SAIHST (Samsung Advanced Institute for Health Sciences and Technology). He is now in charge of Creativity Science Research Institute (CSRI) and Health Mining Research Center (HMRC) as well, all of which are located at Sungkyunkwan University. His recent research interests lie in data mining, health informatics, creativity science, Human-Robot Interaction (HRI), and artificial intelligence techniques in decision making analysis.