

# 텍스트 마이닝 기반의 자산관리 핀테크 기업 핵심 요소 분석: 사용자 리뷰를 바탕으로\*

손애린\*\* · 신왕수\*\*\* · 이준기\*\*\*\*

## 〈목 차〉

I. 서론	3.3 데이터 분석
II. 이론적 배경	IV. 분석결과
2.1 자산관리 핀테크 기업 -뱅크샐러드, 토스	4.1 토픽모델링 결과
2.2 텍스트 마이닝	4.2 속성별 감성 분석 결과
2.2.1 토픽모델링	V. 결론 및 시사점
2.2.2 감성 분석	5.1 연구 결과 토의
2.2.3 속성별 감성 분석	5.2 연구의 한계 및 향후 연구 방향
III. 연구 방법	5.3 연구의 시사점
3.1 데이터 수집	참고문헌
3.2 데이터 전처리	<Abstract>

## I. 서론

자산관리 애플리케이션은 기본적으로 계좌 통합조회 서비스를 제공하며, 이를 통해 가계부를 자동으로 작성하고 소비패턴을 분석하여 맞춤형 금융상품을 추천하는 등 개인 금융 데이터를 활용한 다양한 부가서비스를 제공한다. 이러한 서비스를 활용하여 모바일로 자산을 관리

하는 사람들이 늘고 있으며, 이에 따라 가계부, 카드 사용내역 관리, 세금 계산 등이 가능한 애플리케이션이 인기를 끌며 이른바 ‘애플테크’(애플리케이션과 재테크의 합성어)라는 신조어가 등장하기도 했다(조선비즈, 2017). 게다가 지난 8월 데이터 3법(개인정보보호법·신용정보법·정보통신망법) 개정안이 시행되면서 기업 간의 데이터 거래가 가능해져 빅테크, 은행업, 핀테크

\* 이 연구는 연세대학교 미래융합연구원(ICONS)의 지원으로 수행되었음.

\*\* 연세대학교 정보대학원, [aelin@yonsei.ac.kr](mailto:aelin@yonsei.ac.kr)(주저자)

\*\*\* 연세대학교 정보대학원, [shin.wangsoo@yonsei.ac.kr](mailto:shin.wangsoo@yonsei.ac.kr)

\*\*\*\* 연세대학교 정보대학원, [zlee@yonsei.ac.kr](mailto:zlee@yonsei.ac.kr)(교신저자)

크 업체가 함께 경쟁하게 되어 자산관리 서비스를 제공하던 핀테크 업체에 큰 변화가 있을 것으로 전망된다. 해외에서는 이미 자산관리 애플리케이션이 필수 서비스로 자리 잡고 있으며, 미국의 대표적인 자산관리 애플리케이션 민트(Mint)는 회원 수 5,000만명이 넘고, 이와 유사한 기업인 요들리(Yodlee)는 약 1억명 이상의 사용자를 보유하고 있다(조선일보, 2020). 이와 동시에 국내에서도 핀테크 분야 스타트업인뱅크샐러드와 토스가 비약적인 성장을 하고 있다. 토스는 월간 활성 사용자가 1,000만 명을 돌파했고(매일경제, 2019), 뱅크샐러드는 2017년 6월 애플리케이션 런칭 이후 매달 30%씩 성장하며 대세 핀테크 서비스로 자리매김하였다(중앙일보, 2019). 이렇듯 핀테크 자산관리 서비스가 주목을 받고 있으나 학문적 연구는 미비한 실정이다. 기존 연구들은 핀테크 서비스 모델 연구(서문석, 김동호, 2019), 핀테크 애플리케이션 사용자 경험 연구(정승재, 김승인, 2020; 김현우, 김승인, 2020) 등이 있었으며 실험이나 설문조사에 기반한 연구가 주를 이루었다. 또한 자산관리 애플리케이션에 대한 연구는 이뤄지지 않았다.

따라서 본 연구에서는 앞으로 큰 성장이 기대되는 자산관리 핀테크 기업들의 사용자 리뷰를 분석하여 그동안의 제공되었던 서비스의 장점 및 보완점을 도출하고 자산관리 핀테크 기업의 핵심 요소를 분석하고자 한다. 본 연구에 온라인 리뷰를 활용한 이유는 실험이나 설문조사와 비교하면 비교적 많은 사용자의 솔직한 의견을 반영할 수 있기 때문이며(황해정 등, 2016), 최근 연구에서 대량의 텍스트 데이터를 활용하여 서비스나 제품의 사용 요인을 밝히는

데 텍스트 마이닝 기법을 활용하고 있어 이를 분석 방법으로 적용하였다(이경상 등, 2018; 정폴잎 등, 2019). 구체적으로는 텍스트 마이닝 기법 중 토픽모델링과 감성분석, 그리고 두 방법론을 합친 속성별 감성분석을 활용한다. 본 연구는 성장 잠재력이 뛰어난 자산관리 핀테크 시장에서 서비스들이 갖춰야 할 요소를 텍스트 마이닝 기법을 활용하여 사용자 관점에서 분석하였다는 것에 학술적 의의가 있다. 실무적으로는 자산관리 핀테크 선두기업인 뱅크샐러드와 토스의 애플리케이션 서비스에 대한 보완점을 제시하였다는 것에 의의가 있다.

본 연구는 다음과 같이 구성된다. 제Ⅱ장에서는 자산관리 핀테크 기업에 대해 설명하고, 텍스트 마이닝 관련 선행연구를 살펴본다. 제Ⅲ장에서는 본 연구에서 적용한 연구 방법 및 절차를 세부적으로 설명한다. 제Ⅳ장에서는 분석한 연구 결과를 해석하고, 마지막으로 제Ⅴ장에서는 연구의 의의 및 시사점 등을 제시한다.

## Ⅱ. 이론적 배경

### 2.1 자산관리 핀테크 기업 - 뱅크샐러드, 토스

자산관리 애플리케이션은 계좌통합조회 서비스를 기본적으로 제공하면서, 이를 통해 고객을 대신하여 가계부를 자동으로 작성하고 소비 패턴을 분석하여 맞춤형 금융 상품을 추천하는 등 개인 금융 데이터를 활용한 다양한 부가서비스를 제공하는 것을 말한다. 국내 선두 기업으로는 대표적으로 뱅크샐러드와 토스가 있다.

뱅크샐러드는 가계부 서비스부터 지출 분석, 통합 자산 조회, 금융상품 추천, 신용 및 연금 조회 서비스 등을 제공한다. 뱅크샐러드는 2017년 6월 모바일 애플리케이션을 출시하며 본격적인 모바일 자산관리 서비스를 시작하였다. 2020년 6월 기준 누적 다운로드 수는 약 500만, 연동관리금액은 약 150조원이다(뉴스1, 2020). 토스는 2015년 2월 간편송금 서비스 시작으로 금융 현황 조회, 금융 상품 개설, 투자 등 다양한 금융 서비스를 제공하는 종합 금융 플랫폼으로 성장하였다(KOSME 산업분석 보고서 (핀테크), 2019). 2018년 4월부터 계좌통합조회 및 자산관리 서비스를 제공하였다. 2020년 11월 기준 누적 다운로드 수는 약 5,200만 회이고, 누적 송금액은 약 12조 원이다.

자산관리 애플리케이션에 대한 관심과 사용이 증가하면서 관련 연구도 진행되고 있지만, 기존 연구들은 세대별 사용자 경험 연구(정승재, 김승인, 2020; 김현우, 김승인, 2020), 핀테크 서비스 모델 연구(서문석, 김동호, 2019) 등으로 자산관리 애플리케이션에 대한 통합 연구는 이뤄지지 않았다. 또한, 연구 방법은 설문조사나 인터뷰 방법으로 연구되었다. 설문조사나 인터뷰 방법은 실험 참가자가 연구를 의식하여 실제 경험과 다르게 사회적으로 올바른 응답을 하거나, 설문 응답 상황에서 편향(Bias)이 발생할 수 있다(Bhattacharjee, 2012). 즉, 사용자들이 자산관리 애플리케이션을 사용하면서 실제로 느끼는 경험을 연구하는데 한계가 있을 수 있다. 그리하여 본 연구에서는 자산관리 서비스를 제공하는 애플리케이션의 리뷰를 수집 후 분석하여 자산관리 서비스를 사용하는 핵심 요인을 탐구하고자 한다. 온라인 리뷰 분석은 실

협이나 설문조사와 비교하면 비교적 많은 사용자의 솔직한 의견을 반영할 수 있기 때문이다(황해정 등, 2016). 최근 연구에서 서비스나 제품의 요인을 밝히는 데 텍스트 마이닝 기법을 활용하고 있다(이경상 등, 2018; 정폴잎 등, 2019).

## 2.2 텍스트 마이닝

### 2.2.1 토픽모델링

토픽모델링은 구조화되지 않은 문헌에서 지정한 수의 잠재적인 토픽을 분류해내는 확률 분포 모델이다(Blei, 2012). 토픽모델링에 쓰이는 LDA 알고리즘은 잠재적인 주제들이 무작위로 섞여 문서를 구성하고 있다고 가정하며, 디리클레 확률 분포를 이용하여 특정 단어가 주제에 포함될 확률을 계산한다(Blei et al., 2003). 토픽모델링은 해당 문헌을 구성하는 토픽의 수를 연구자가 직접 지정해야 하며, 지정해주는 토픽의 수에 따라 결과가 달라질 수 있다. 따라서 분석 결과에 객관성을 확보하기 위해 일반적으로 적정한 토픽의 수는 혼잡도(Perplexity)와 일관성도(Coherence)를 통해 평가한다. 혼잡도는 토픽모델링의 결과가 문헌의 내용을 얼마나 잘 반영하는지를 나타내는 수치로, 값이 낮을수록 좋은 결과이며(Chang et al., 2009), 일관성도는 토픽 내 키워드의 유사도를 분석한 수치로 높을수록 좋은 결과이다(Newman et al., 2010). 최근에는 이러한 토픽 모델링을 다른 분석 기법과 함께 활용하는 연구가 이루어지고 있다(최병설, 김남규, 2019; 홍태호 등, 2018).

### 2.2.2 감성 분석

감성 분석은 텍스트를 통해 표현된 의견이나 감정, 태도를 계산적으로 연구하는 학문이다 (Medhat et al., 2014). 소셜미디어의 확산으로 인해 관심이 증가하였으며, 기업과 조직은 감성 분석을 통해 사용자와 직원들이 자사의 제품과 서비스를 어떻게 느끼는지에 대해 알고 싶어했다 (Appel et al., 2015). 그로 인해 감성 분석은 2000년대 초부터 자연어 처리 분야에서 가장 활발한 연구 분야로 성장했다 (Zhang et al., 2018).

감성 분석의 목표는 어떤 주제에 대한 사람들의 의견을 결정하는 것이다 (Schouten et al., 2015). 구조적으로는 주로 문서 수준, 문장 수준, 양상 수준 등 세 가지 수준으로 세분화되며, 그 결과를 긍정, 부정, 중립으로 분류한다 (Zhang et al., 2018). 감성 분석의 연구방향은 크게 사전기반 감성 분석과 머신러닝 기반 감성 분석으로 나뉘며, 머신러닝 기반 감성 분석에서는 지도학습과 비지도학습으로 다시 나뉜다. 사전기반 감성 분석은 준비된 감성사전을 기반으로 각 어휘의 가중치를 측정해 감성분류를 하는 방식이다. 지도학습 기반 감성 분석은 데이터로부터 많은 특징을 습득한다 (Schouten et al., 2015). 비지도학습 기반 감성 분석은 감성 어휘, 문법 분석, 통사적 패턴 등을 활용하는 다양한 방법이 포함되었다. 딥러닝의 부상은 많은 연구 분야에 변화를 가져왔고, 감성 분석을 포함하는 NLP 분야도 딥러닝을 통해 많은 성과 개선을 이루었다. 감성 분석 분야에도 최근에는 딥러닝을 적용하여 높은 성능을 보이고 있다 (Zhang et al., 2018).

### 2.2.3 속성별 감성 분석

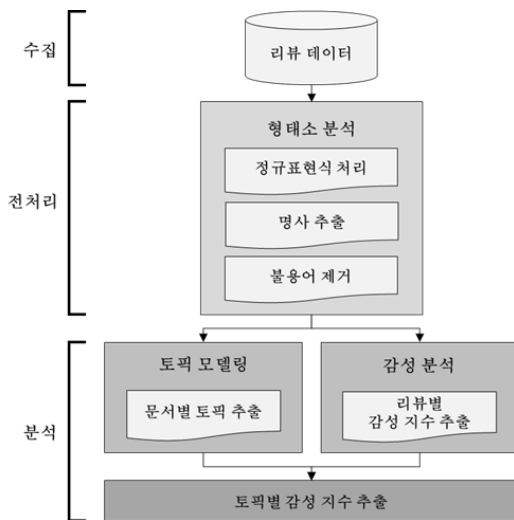
속성별 감성 분석 (Aspect-Sentiment Analysis)은 문서 수준, 문장 수준을 넘어 제품이나 서비스의 특정 측면에 대해 사람들의 의견을 추출하고 분석하는 것이다. 예를 들면, 제품 리뷰에서 제품에 대한 전반적인 감정은 긍정적이지만, 제품의 특정 측면에 대한 의견은 부정적일 수 있다 (Zhang et al., 2018). 속성별 감성 분석에서는 주제 추출, 실체 추출, 속성별 감성 분류와 같은 하위 작업으로 구성된다. 예를 들어 ‘맥북은 성능은 좋지만 가격이 비싸다.’라는 문장에서 ‘맥북’은 실체, ‘성능’과 ‘가격’을 주제로 추출해야 한다 (Zhang et al., 2018). 따라서 주제를 추출하는 Aspect-Detection 과정과 주제에 따른 감성 분석의 과정으로 이루어진다 (Schouten et al., 2015).

Aspect-Detection은 빈도 기반, 문맥 기반, 지도학습 머신러닝 기반, 비지도 학습 머신러닝 기반, 하이브리드 기반의 접근이 있다. 먼저 빈도 기반 접근은 가장 간단한 방법으로 출현 빈도가 높은 단어를 주제로 추출한다. 다만 빈도 기반 방법은 사람들이 명시적으로 언급하지 않는 주제는 선정되지 않을 확률이 높다. 문맥 기반 접근은 빈도를 고려하는 대신에 문맥의 관계를 통해 주제를 추출하는 방법이다. 지도학습 머신러닝 기반 접근에서 순수하게 지도학습을 통한 머신러닝 방법으로 주제를 추출하는 방법은 많지 않으며, 그 이유는 학습을 위한 특성이 주로 다른 방법론을 통해 이루어지기 때문이다. 또한 라벨링 기준도 굉장히 까다로운 부분이다. 따라서 Aspect-Detection에서 비지도 학습 머신러닝 기반 접근이 가장 활발하게 활용되고 있으며, 그 중에서도 LDA 토픽모델링이 가장 많

이 활용되고 있다(Schouten et al., 2015; 홍태호 et al., 2018).

### Ⅲ. 연구 방법

제 3장에서는 본 연구의 연구 방법에 대해 세부적으로 설명한다. 연구 대상인 두 서비스는 기능과 서비스 기간에서 다소 차이가 존재하므로 두 서비스를 독립적으로 분석하였다. 전체적인 연구 절차는 데이터 수집, 데이터 전처리, 데이터 분석의 과정으로 이루어지며, 이를 설명하는 모형도는 <그림1>과 같다.



<그림 1>

#### 3.1 데이터 수집

데이터 수집 단계에서는 파이썬 기반의 웹 크롤링 패키지 Selenium을 통해 구글 플레이 스토어의 ‘뱅크샐러드’와 ‘토스’ 애플리케이션

리뷰를 각각 수집하였다. 수집한뱅크샐러드 리뷰의 작성 기간은 2017년 6월부터 2020년 10월 까지이며 데이터 수는 총 9,924건이었다. 수집한 토스 리뷰의 작성 기간은 2018년 4월부터 2020년 10월까지이며 데이터 수는 총 33,356건이었다. 수집 기간은 각각 두 서비스가 자산 관리 서비스를 시작한 시점으로 하였다.

#### 3.2 데이터 전처리

데이터 전처리 단계에서는 정규표현식 처리를 통하여 특수문자를 제거하고 한글, 알파벳, 숫자만 보존하였다. 다음으로 LDA토픽모델링을 위해 KoNLPy 패키지에서 제공하는 Okt 형태소 분석기를 사용하여 토큰화 작업을 실시하였다. 마지막으로 분석의 성능을 낮추는 ‘자주’, ‘이후’, ‘갑자기’, ‘자꾸’ 등의 불용어를 제거하였다. 또한 지나치게 짧은 리뷰는 토픽 할당에 정확도를 떨어뜨릴 수 있으므로, 토큰 개수가 5개 미만인 리뷰는 제거하였다. 그 결과 최종적으로뱅크샐러드 5,529건, 토스 9,441건을 분석 대상으로 추출하였다.

#### 3.3 데이터 분석

데이터 분석 단계에서는 LDA토픽모델링과 속성별 감성 분석의 순서로 진행하였다. LDA토픽모델링은 토픽모델링 기능을 제공하는 Gensim 패키지를 사용하였다. 토픽의 수를 정하는 기준은 Gensim 패키지에서 제공하는 Coherence 값과 LDA토픽모델링 결과를 시각화 하여 보여주는 pyLDAvis 값을 통해 선정하였다.

다음으로 LDA토픽모델링 결과를 활용하여 속성별 감성 분석을 실시하였다. 주제의 중복을 방지하기 위해 LDA토픽모델링 결과에서 각 주제가 할당될 확률을 추출하여 특정 주제가 할당될 확률이 과반(0.5) 이상의 값을 나타내는 리뷰만을 감성 분석의 대상으로 삼았다. 그 결과뱅크샐러드는 3,763건, 토스는 6,196건을 속성별 감성 분석 대상으로 선정하였다. 감성 분석의 결과는 각 리뷰에 대한 소비자의 긍정/부정 여부와 그 정도를 부여하였다. 긍정/부정의 정도는 0~1 사이의 값으로 나타냈으며, 1에 가까울수록 긍정(positive), 0에 가까울수록 부정(negative), 0.3~0.7은 중립(neutral)으로 산출하였다. 감성 분석은 BERT기반의 사전학습모델을 활용하는 TEANAPS를 활용하였다. TEANAPS는 연구 목적으로 제작된 python 언어 기반의 오픈 소스 라이브러리이다.

## IV. 분석 결과

### 4.1 토픽모델링 결과

토픽모델링에서 가장 중요한 하이퍼 파라미터는 토픽의 수이다. 객관적인 토픽의 수를 선정하기 위해서 Gensim 패키지에서 제공하는 Coherence 값과 pyLDAvis에서 제공하는 토픽별 군집 결과를 교차로 확인하였다. 토픽의 범위는 동일하게 5~10개로 측정하였다. 기본적으로 Coherence값이 급격하게 높아지는 변곡점의 값을 선정하고, pyLDAvis를 통해 토픽별 군집 결과가 겹치지 않는 값으로 선정하였다. 그 결과뱅크샐러드의 토픽 수는 Coherence값

(0.513)을 나타낸 7개, 토스의 토픽 수는 Coherence값(0.447)을 나타낸 7개로 선정하였다. 또한 각 토픽의 토픽명은 함께 추출된 키워드의 구성을 고려하여 3명의 연구자가 함께 협의하여 최종적으로 도출하였다.

뱅크샐러드는 총 7개의 토픽을 도출하였다. 먼저 ‘가계부’ 토픽은 ‘지출’, ‘가계부’, ‘자산’, ‘관리’, ‘수입’의 키워드를 통해 지출과 수입 등의 내역을 정리하여 보여주는 가계부 기능에 관한 토픽으로 해석하였다. ‘금융사 추가’ 토픽은 ‘카카오뱅크’, ‘연동’, ‘추가’, ‘증권’, ‘보험’의 키워드가 도출되었고, 이를 통해 은행이나 보험사 등의 자산 연동을 위한 금융사 추가에 관한 토픽으로 해석하였다. ‘사용자 설정’은 ‘사용’, ‘설정’, ‘예산’, ‘계좌’, ‘카테고리’의 키워드가 추출되었다. 이는 사용자가 뱅크샐러드를 사용할 때 예산이나 지출 카테고리를 임의로 설정하는 사용자 설정에 관한 토픽으로 해석하였다. ‘업데이트’ 토픽은 ‘업데이트’, ‘인증’, ‘사용’, ‘실행’, ‘동기화’의 키워드로 추출되었는데 이는 애플리케이션의 기능 업데이트나 앱 사용시 자산 내역 동기화에 관한 토픽으로 해석하였다. ‘자산 등록’ 토픽은 ‘연동’, ‘오류’, ‘공인인증서’, ‘은행’, ‘등록’의 키워드가 추출되었다. 이는 금융 자산을 등록할 때 생겨나는 사용자 경험에 관한 토픽으로 해석하였다. ‘자산 정보’ 토픽은 ‘금융’, ‘정보’, ‘자산’, ‘대출’, ‘관리’로 키워드가 추출되었으며, 이는 사용자의 자산에 대한 토픽으로 해석하였다. 마지막으로 ‘카드 사용’ 토픽은 ‘카드’, ‘사용’, ‘결제’, ‘금액’, ‘신용카드’ 키워드가 추출되었는데, 이를 통해 사용자들이 사용한 카드에 대한 정보를 표현하는 토픽으로 해석하였다.

토스 또한 총 7개의 토픽을 도출하였다. 먼저 ‘가입’ 토픽은 ‘인증’, ‘가입’, ‘번호’, ‘본인’, ‘명의’ 키워드가 추출되었다. 이를 통해 서비스 가입에 대한 토픽으로 해석하였다. ‘광고 이벤트’ 토픽이 추출되었는데, 키워드는 ‘사람’, ‘사용’, ‘친구’, ‘광고’, ‘탈퇴’로 추출되었다. 토스가 진행하는 친구 추천 이벤트나 다양한 광고에 대한 토픽으로 해석하였다. 다음으로 ‘사용 오류’ 토픽은 ‘오류’, ‘확인’, ‘계좌’, ‘잔액’, ‘로그인’으로 키워드가 추출되었다. 해당 토픽은 서비스 사용시에 발생하는 각종 오류에 대한

토픽으로 해석하였다. ‘송금’ 토픽은 ‘송금’, ‘계좌’, ‘수수료’, ‘이체’, ‘무료’ 키워드가 추출되었는데, 이는 토스의 송금 기능에 관한 토픽으로 해석하였다. 다음으로뱅크샐러드와 마찬가지로 ‘업데이트’ 토픽이다. ‘업데이트’, ‘알림’, ‘사용’, ‘화면’, ‘버전’의 키워드가 추출되었으며 애플리케이션의 업데이트에 관한 토픽으로 해석하였다. ‘자산 등록’ 토픽은 ‘계좌’, ‘등록’, ‘카드’, ‘인증서’, ‘입력’의 키워드가 추출되었다. 이를 통해 서비스를 사용할 때 계좌를 등록하는 것에 대한 토픽으로 해석하였다.

<표 1> 뱅크샐러드와 토스의 토픽모델링 결과

서비스	토픽명	토픽 키워드	토픽 해석
뱅크샐러드	가계부	지출, 가계부, 자산, 관리, 수입	지출과 수입을 보여주는 가계부 기능에 관한 토픽
	금융사 추가	카카오뱅크, 연동, 추가, 증권, 보험	금융사 연동을 위한 금융사 추가 요청에 관한 토픽
	사용자 설정	사용, 설정, 예산, 계좌, 카테고리	예산 및 지출 카테고리 설정에 관한 토픽
	업데이트	업데이트, 인증, 사용, 실행, 동기화	애플리케이션 기능 업데이트와 동기화에 관한 토픽
	자산 등록	연동, 오류, 공인인증서, 은행, 등록	금융 자산 등록에 관한 토픽
	자산 정보	금융, 정보, 자산, 대출, 관리	금융 자산을 보여주는 기능에 관한 토픽
	카드 사용	카드, 사용, 결제, 금액, 신용카드	지출 및 카드 사용에 관한 토픽
토스	가입	인증, 가입, 번호, 본인, 명의	가입에 관한 토픽
	광고 이벤트	사람, 사용, 친구, 광고, 탈퇴	광고 및 이벤트에 관한 토픽
	사용 오류	오류, 확인, 계좌, 잔액, 로그인	사용 오류에 관한 토픽
	송금	송금, 계좌, 수수료, 이체, 무료	송금 기능에 관한 토픽
	업데이트	업데이트, 알림, 사용, 화면, 버전	애플리케이션 기능 업데이트와 동기화에 관한 토픽
	자산 등록	계좌, 등록, 카드, 인증서, 입력	금융 자산 등록에 관한 토픽
	투자	투자, 고객, 문의, 소액, 원금	투자 기능에 관한 토픽

마지막으로 ‘투자’ 토픽인데, 키워드는 ‘투자’, ‘고객’, ‘문의’, ‘소액’, ‘원금’이 추출되었다. 토스에서 꾸준히 제공해오던 자산 투자에 대한 토픽으로 해석할 수 있었다. 전체적인 토픽모델링 결과를 정리하면 <표1>과 같다.

#### 4.2 속성별 감성 분석 결과

속성별 감성 분석은 LDA토픽모델링을 통해 도출된 각각의 토픽별로 감성지수를 산출하였다. 이를 통해 어떤 요소에서 긍정/부정의 경향이 나타나는지 분석하였다.

그 결과뱅크샐러드의 전체 리뷰에 대한 감성 지수는 0.549로 도출되었으며, 각 속성별 감성 지수는 가계부(0.631), 금융사 추가(0.726), 사용자 설정(0.524), 업데이트(0.347), 자산 등록(0.414), 자산 정보(0.673), 카드 사용(0.518)로 나타났다. 총 3,763개의 리뷰 중 1,895개가

긍정적인 리뷰로 나타났으며, 1,533개는 부정적인 리뷰로 나타났다.

토스의 전체 리뷰에 대한 감성 지수는 0.398로 나타났으며, 각 속성별 감성 지수는 가입(0.335), 광고 이벤트(0.340), 사용 오류(0.264), 송금(0.561), 업데이트(0.390), 자산 등록(0.378), 투자(0.319)로 나타났다. 전체 리뷰 6,196개 중 긍정적인 리뷰는 2,083개, 부정적인 리뷰는 3,572개로 도출되었다. 전체적인 결과를 정리하면 <표 2>와 같다.

## V. 결론 및 시사점

### 5.1 연구 결과 토의

본 연구는 국내 자산관리 핀테크 기업이 갖춰야 할 핵심 요소가 무엇인지 텍스트 마이닝

<표 2>뱅크샐러드와 토스의 속성별 감성 분석 결과

서비스	토픽명	속성별 감성 지수	전체 문서 수	긍정 문서 수	부정 문서 수	중립 문서 수
뱅크샐러드	가계부	0.631	612	373	193	46
	금융사 추가	0.726	800	568	158	74
	사용자 설정	0.524	235	110	105	20
	업데이트	0.347	566	148	361	57
	자산 등록	0.414	773	271	440	62
	자산 정보	0.673	373	243	104	26
	카드 사용	0.518	404	182	172	50
	<b>전체</b>	<b>0.549</b>	<b>3,763</b>	<b>1,895</b>	<b>1,533</b>	<b>335</b>
토스	가입	0.335	869	208	547	114
	광고 이벤트	0.340	633	183	406	44
	사용 오류	0.264	867	158	637	72
	송금	0.561	1,634	880	658	96
	업데이트	0.390	607	194	354	59
	자산 등록	0.378	1,224	359	722	143
	투자	0.319	362	101	248	13
	<b>전체</b>	<b>0.398</b>	<b>6,196</b>	<b>2,083</b>	<b>3,572</b>	<b>541</b>



기법을 활용하여 파악하고자 하였다. 이를 위해 자산관리 애플리케이션 리뷰를 분석하였다. LDA토픽모델링을 통해 주요 요소를 추출하고, 속성별 감성 분석을 통해 사용자들이 자산관리 서비스의 어떤 요소들에 대해서 주로 반응을 보이며, 그것이 긍정적인 반응인지, 부정적인 반응인지를 살펴보았다.

토픽모델링 분석 결과를 살펴보면뱅크샐러드는 상대적으로 ‘가계부’, ‘사용자 설정’, ‘자산 정보’, ‘카드 사용’ 등의 토픽을 통해 자산관리 서비스의 기능 위주의 토픽이 주로 추출되었다는 것을 알 수 있었다. 반대로 토스는 ‘가입’, ‘사용 오류’, ‘자산 등록’ 등의 토픽을 통해 자산관리 서비스를 사용하는 관점에서의 토픽이 주로 추출되었다는 것을 알 수 있었다.뱅크샐러드에서 추출된 ‘가계부’, ‘사용자 설정’, ‘자산 정보’, ‘카드 사용’ 등의 토픽은 사용자들에게 자산관리 서비스의 정체성이 잘 전달되고 있다는 것을 확인할 수 있는 부분이었다. 토스는 ‘가입’, ‘사용 오류’, ‘자산 등록’ 등의 토픽을 통해 애플리케이션의 사용 측면에 대해 사용자들이 많은 언급을 하고 있다는 것을 알 수 있었다. 두 서비스에서 동시에 출현한 ‘자산 등록’ 토픽은 흩어져 있는 금융 자산들을 연동하여 한 눈에 보여주는 기능에서 사용자들의 경험이 잘 발현되고 있다는 것을 알 수 있었다. 주목할 점은 ‘금융사 추가’라는 토픽이뱅크샐러드에서만 추출되었는데, 이는 더 많은 금융사들을 연동하여 한 눈에 보고 싶어하는 사용자들의 욕구가 발현된 토픽이라고 볼 수 있다. 두 서비스에서 동시에 추출된 또 다른 토픽은 ‘업데이트’라는 토픽이었다. 자산관리 서비스는 기능의 특성상 애플리케이션에 접속할 때마다

모든 자산을 동기화 시켜서 업데이트 된 자산 내역을 보여주어야 하는데, 많은 서버로부터 동시에 많은 데이터를 끌어와야 하기 때문에 네트워크 상황과 디바이스의 사양에 따라 다양한 이슈가 발생할 수 있다. 여기서 더 나아가 모바일 애플리케이션 서비스라는 특성상 지속적으로 서비스를 최적화하고 기능을 추가해야 하기 때문에 애플리케이션의 업데이트를 자주 하는 경우가 많다. 따라서 ‘업데이트’라는 토픽은 자산관리 서비스 애플리케이션의 안정성에 대한 사용자들의 욕구가 있다는 점을 보여주는 토픽이었다. 마지막으로 토스에서만 추출된 ‘광고 이벤트’라는 토픽이 있었는데, 이는 토스가 장기간 진행하였던 각종 가입 이벤트에 관련된 토픽이었다. 토스는 가입자들을 늘리기 위해 친구를 초대하여 사용자의 사용에 대한 리워드를 제공하며 애플리케이션의 성격을 다소 보여주었는데, 이러한 측면 또한 사용자에게 매우 의미있게 다가가고 있다는 것을 보여주는 토픽이었다.

이러한 토픽 분류를 활용하여뱅크샐러드와 토스의 속성별 감성 분석을 진행하였다. 그 결과뱅크샐러드의 속성 중 감성 지수가 가장 높게 나타난 속성은 ‘금융사 추가(0.726)’ 토픽이었다. ‘금융사 추가’ 토픽의 감성 지수가 가장 높게 나타난 것은 사용자들이 많은 금융사에 흩어져 있는 자산을 종합해서 보여 주는뱅크샐러드의 자산관리 서비스가 긍정적으로 작용하고 있다는 것을 알 수 있는 대목이었다. 토픽의 키워드에서 볼 수 있듯이 대부분 ‘카카오뱅크’ 등의 인터넷 뱅킹에 대한 추가 요청이 주를 이루었으며, 그와 동시에 기존에 제공하고 있는 기능에 만족하고 있다는 것을 알 수 있었다. 반대로 가장 낮은 감성 지수를 나타낸 주제는 ‘업

데이트(0.347)'였다. 자산관리 기능을 제공하는 것은 사용자들에게 긍정적으로 작용하고 있지만, 사용자들이 애플리케이션을 사용할 때마다 자산 내역을 동기화하기 위해서 많은 시간을 기다려야 하는 부분에 대해서는 불만족하고 있다는 것을 알 수 있었다. 또한 서비스 최적화와 기능 개선을 위한 애플리케이션 업데이트가 잦을수록 사용자들은 오히려 부정적인 반응을 보인다는 것을 알 수 있었다. 이는 새로운 것을 익히거나 변화에 적응해야 하는 것을 사용자들이 부정적으로 받아들이고 있다는 점을 알 수 있는 부분이었다. 토스의 감성 지수 수치(0.398)는 전체적으로뱅크샐러드의 감성 지수 수치(0.549)보다 매우 낮게 나타났다. 가장 높은 감성 지수를 기록한 '송금(0.561)' 토픽이 유일하게 0.5를 넘었으며, 가장 낮은 감성 지수를 기록한 '사용 오류(0.264)' 토픽은 0.3보다 낮은 수치를 보였다. 전체적으로 토스의 사용자 경험이뱅크샐러드보다 매우 부정적으로 작용하고 있음을 알 수 있다. 그러나 추출된 토픽의 분포를 살펴보면 다소 차이가 있다는 점을 알 수 있었다. 토스는뱅크샐러드와는 다르게 기능과 관련된 토픽보다는 사용과 관련된 토픽이 주를 이루었고, 두 서비스에서 동일하게 추출된 '업데이트'와 '자산 등록'에 대한 토픽은 각각(0.347, 0.390), (0.414, 0.378)로 큰 차이를 보이지 않았다. 이러한 점은 자산관리 서비스의 사용과 관련된 측면이 기능적인 측면보다 사용자들에게 부정적으로 발현되고 있다고 해석할 수 있는 부분이었다. 이를 뒷받침하듯, 애플리케이션의 사용과 관련된 토스의 '가입(0.335)', '사용 오류(0.264)', '자산 등록(0.378)' 토픽과뱅크샐러드의 '사용자 설정(0.524)', '자산 등록

(0.414)' 토픽에서는 다른 토픽들보다 부정적인 수치를 볼 수 있었다. 결국 사용자들은 자산관리 서비스를 사용함에 있어서 기존의 서비스들이 제공하던 사용자 경험과 유사한 모습, 즉 서비스를 시작하고 사용하는 데 많은 장애가 존재하거나, 많은 단계를 거쳐야 하는 부분들에 대해 유독 부정적으로 반응한다는 것을 알 수 있는 부분이었다. 토스에서 추출된 토픽 중 유일하게 높은 감성 지수를 기록한 토픽은 '송금(0.561)' 토픽이었으며, 이는 사용자들이 토스의 기능적 측면에서는 다른 토픽보다도 긍정적으로 받아들이고 있다는 것을 알 수 있는 대목이었다.

결과적으로 본 연구는 분석 결과를 통해 다음과 같은 결론을 내릴 수 있었다. 국내 자산관리 핀테크 서비스의 기능적인 요소들은 사용자들에게 긍정적으로 작용하고 있었다. 사용자들은 지금껏 겪어보지 못한 자산관리 서비스의 기능들을 대체로 긍정적으로 받아들이고 있었으며, 더 많은 기능과 편리함을 추구하고 있었다. 반대로 자산관리 서비스의 사용과 관련된 요소에서는 부정적인 감정이 많이 표출되었으며 이는 앞으로 개선해야 할 부분으로 도출되었다. 사용자들은 자산관리 서비스를 사용하기 위해 반드시 거쳐야 할 과정들인 자산 등록이나 사용자 설정 등의 과정에 대해서 더 편리하고 단순한 과정을 원하는 것으로 나타났다. 또한 자산 내역의 동기화를 위한 과정과 애플리케이션의 안정성 측면에서도 부정적인 반응이 도출되며 개선해야 할 부분으로 나타났다. 이러한 부분은 단순히 자산관리 애플리케이션을 운영하는 단일 기업의 영역을 넘어 전 금융권이 협력하며 개선해야 할 부분일 것이다.

## 5.2 연구의 한계 및 향후 연구 방향

본 연구의 한계점과 이를 보완한 향후 연구 방향은 다음과 같다. 첫째, 본 연구는 성장하는 핀테크 산업 속에서 자산관리 핀테크 기업이 갖춰야 할 핵심 요소를 파악하려는 연구였다. 그러나 자산관리 서비스 중에서도뱅크샐러드와 토스만을 분석하였다는 한계점이 존재한다. 현재 국내에서 운영 중인 자산관리 서비스는 두 서비스 외에도 핀크, 카카오페이 등이 있다. 다만 연구 대상을 선정하는 과정에서뱅크샐러드와 토스 외의 서비스는 시행 시점이 1년 미만이거나 수집할 수 있는 리뷰 데이터 수가 5천 건이 되지 않아 분석 대상에서 제외하였다. 따라서 향후 연구에서는뱅크샐러드와 토스 외에도 국내에서 운영 중인 자산관리 서비스 애플리케이션을 모두 종합하여 분석하는 연구가 필요할 것이다.

둘째, 본 연구는 텍스트 마이닝 방법론을 적용한 연구이며, 세부적인 방법으로는 LDA토픽 모델링과 속성별 감성 분석을 실시하였다. LDA토픽 모델링은 다량의 문서에서 주요 토픽을 추출해주는 효과적인 방법이며, 속성별 감성 분석의 Aspect-Detection 단계에서 비지도 학습의 방법으로 널리 사용되고 있는 알고리즘이다. 다만 그 결과의 해석에 있어서 연구자의 주관적 개입된다는 한계점이 존재한다. 본 연구에서는 이러한 한계점을 극복하기 위해 토픽 개수를 선정하는 과정에서 Coherence 수치와 토픽별 군집 시각화 결과를 교차로 확인하여 선정하였으며, 토픽명 선정 과정에서도 다수의 연구자가 협의하여 선정하였다. 그러나 본질적인 방법론의 한계점을 해결하기 위해서는 조금 더

객관성을 보장할 수 있는 방법론을 적용하는 것이 필요할 것이다.

마지막으로 본 연구는 분석을 위해 구글 플레이 스토어의 애플리케이션 리뷰 데이터를 수집하였다. 연구 대상을 자산관리 서비스 애플리케이션으로 한정 지어 자료를 수집하였으나, 두 서비스 모두 자산관리 서비스와 관련된 리뷰 외에도 다른 유형의 리뷰가 다소 존재하였다. 또한 하나의 리뷰 안에서 다수의 의견을 이야기하는 리뷰도 존재하였다. 이는 하나의 애플리케이션이 자산관리 외에도 다양한 서비스를 제공하고 있으며, 서비스 외에도 다양한 마케팅을 진행하고 있기 때문에 발생하는 문제였다. 이러한 문제를 해결하기 위해 LDA토픽 모델링 결과에서 하나의 토픽 할당 확률이 0.5 이상인 리뷰만을 감성 분석의 대상으로 선정하여 최대한 노이즈를 제거하였다. 향후 연구에서는 전처리 단계에서 이러한 한계점을 해결할 수 있는 방법을 적용하여 분석을 진행해야 할 것이다.

## 5.3 연구의 시사점

본 연구의 학술적 시사점은 다음과 같다. 첫째, 성장 잠재력이 뛰어난 자산관리 핀테크 시장에서 서비스들이 갖춰야 할 요소를 사용자의 관점에서 분석하였다. 이를 통해 국내 자산관리 핀테크 기업이 성공적으로 자리 잡기 위해 어떤 요소를 갖춰야 하는지 도출해낼 수 있었다. 국내는 아직 자산관리 핀테크 관련 연구가 부족한 실정이었다. 특히나 텍스트 마이닝 기법을 통해 분석한 연구는 전무하였다. 본 연구는 온라인 리뷰 데이터를 활용하여 텍스트 마이닝 기법으로 분석함으로써 사용자들이 어떤 요소

에 긍정적인 반응을 보이며, 어떤 요소에 대해 부정적인 반응을 가지고 있는지 파악할 수 있었다.

둘째, 본 연구는 핀테크 자산관리 애플리케이션의 리뷰 데이터를 기반으로 텍스트 마이닝 기법을 적용하였으며, LDA토픽모델링과 감성 분석을 연결한 속성별 감성 분석 방법론을 적용하여 분석하였다. 기존에 수행되었던 핀테크 자산관리 서비스에 대한 연구는 정성적 연구를 비롯한 인터뷰, 설문조사 등의 방법을 통해 연구가 진행되었다. 그러나 본 연구는 온라인 리뷰 데이터를 텍스트 마이닝 기법을 활용해 분석 함으로서 다수의 사용자들의 솔직한 의견을 대량으로 분석하여 연구 결과를 도출하였다는 의의가 있다. 또한 방법론 측면에서 LDA토픽모델링과 감성 분석을 연결하여 애플리케이션 리뷰를 각 속성별로 분리해 감성 분석을 실시하였다는데 의의가 있다. 기존의 연구들은 단순히 토픽모델링을 통해서 주요 토픽만을 추출하거나, 문서 단위나 리뷰 단위에서의 감성 분석만을 진행하였다. 본 연구는 LDA토픽모델링을 통해 주요 토픽을 추출하고, 해당 토픽들의 감성 지수를 추출하였다. 이를 통해 연구 대상 서비스의 주요 요소별 감성 지수를 분리해서 살펴볼 수 있었다.

마지막으로 실무적 시사점으로는 본 연구가 핀테크 자산관리 서비스의 선두 기업인뱅크샐러드와 토스의 애플리케이션에 대한 보완점을 제시하였다는 점이다. 본 연구는 애플리케이션에 대한 온라인 리뷰 텍스트 데이터를 분석하여 사용자들이 어떤 요소를 주로 언급하는지, 나아가 어떤 요소에 대해 더 부정적으로 받아들이고 있는지 분석 결과를 통해 제시하였다.

그 결과 사용자들이 자산관리 서비스의 기능적인 측면에서는 긍정적으로 받아들이고 있지만, 사용에 대해서는 여전히 많은 불편함과 함께 부정적인 감정을 느끼고 있다는 것을 알 수 있었다. 이러한 연구 결과를 통해뱅크샐러드와 토스는 자사의 애플리케이션 서비스의 요소 중에서 어떤 부분을 먼저 개선하여 사용자들의 긍정적인 경험을 이끌어낼 수 있을지에 대해 참고할 수 있을 것이다. 결론적으로 본 연구는 방대한 온라인 리뷰 텍스트를 통해서 국내 핀테크 자산관리 서비스의 개선과 사용자 만족도 향상에 기여 할 수 있다.

## 참고문헌

- 김현우, 김승인. “핀테크 애플리케이션의 사용자 경험 연구-토스와 카카오뱅크를 중심으로.” 디지털융복합연구 18.1, pp.287-293.
- 서문석, 김동호. “빅데이터 기반의 핀테크서비스 모델의 변화 방향에 관한 연구.” e-비즈니스연구 20.2, 2019, pp.195-213.
- 이경상, 김기윤, 이준기. “스마트 모빌리티의 구매요인 분석: Mental Accounting Theory 와 텍스트 마이닝을 중심으로.” 대한경영학회지 31.11, 2018, pp.2147-2168.
- 정승재, 김승인. “Z 세대의 모바일 핀테크 서비스 사용자 경험 연구-카카오페이와 토스를 중심으로.” 디지털융복합연구 18.1, 2020, pp.315-320.
- 정폴잎, 안현철, 곽기영. “텍스트 마이닝과 소셜

- 네트워크 분석을 이용한 스마트폰 디자인의 핵심속성 및 가치 식별.” 대한경영학회지 32.1, 2019, pp.27-47.
- 최병설, 김남규. “감정 딥러닝 필터를 활용한 토픽 모델링 방법론.” 정보시스템연구 28.4, 2019, pp.271-291.
- 홍태호, 임강, 박지영. “LDA 를 이용한 온라인 리뷰의 다중 토픽별 감성분석 -TripAdvisor 사례를 중심으로.” 정보시스템연구 27.1, 2018, pp.89-110.
- 황해정, 심혜린, 최준호. “빅데이터 분석을 활용한 사용자 경험 평가 방법론 탐색 : 아마존 에코에 대한 온라인 리뷰 분석을 중심으로”, 한국콘텐츠학회논문지, 제 16권, 제8호, 2016, pp.517-528.
- KOSME 기업심사센터, “KOSME 산업분석 Report, ICT: 핀테크” 융합금융처 산업 Report 2019-9호, 2019.
- Appel, Orestes, Francisco Chiclana, and Jenny Carter. “Main concepts, state of the art and future research questions in sentiment analysis.” *Acta Polytechnica Hungarica* 12.3, 2015, pp. 87-108.
- Bhattacharjee, Anol, “Social Science Research: Principles, Methods, and Practices” *Textbooks Collection*. 3, 2012.
- Blei, David M. “Probabilistic topic models.” *Communications of the ACM* 55.4, 2012, pp. 77-84.
- Blei, David M. Andrew Y. Ng, and Michael I. Jordan. “Latent dirichlet allocation.” *Journal of machine Learning research* 3. Jan, 2003, pp. 993-1022.
- Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J., and Blei, D. “Reading tea leaves: How humans interpret topic models.” *Advances in neural information processing systems* 22, 2009, pp. 288-296.
- Medhat, Walaa, Ahmed Hassan, and Hoda Korashy. “Sentiment analysis algorithms and applications: A survey.” *Ain Shams engineering journal* 5.4, 2014, pp. 1093-1113.
- Newman, D., Lau, J. H., Grieser, K., and Baldwin, T. “Automatic evaluation of topic coherence.” *Human language technologies: The 2010 annual conference of the North American chapter of the association for computational linguistics*. 2010.
- Schouten, Kim, and Flavius Frasinicar. “Survey on aspect-level sentiment analysis.” *IEEE Transactions on Knowledge and Data Engineering* 28.3, 2015 pp. 813-830.
- Zhang, Lei, Shuai Wang, and Bing Liu. “Deep learning for sentiment analysis: A survey.” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 8.4 (2018): e1253.
- 뉴스1, “[NFF2020]뱅크샐러드 김태훈 금융업의 기회, 디지털화가 관건”, 송화연, 최동현 기고, 2020.07.16.
- 매일경제, “토스 vs 뱅샐 1등 금융 플랫폼 전쟁... 토스, 결제·송금 기반 서비스로

인터넷은행 도전뱅크샐러드는 데이터 기반 서비스로 소비자 눈길”, 김강래 기고, 2019.10.30.

조선비즈, “모바일로 자산관리하는 시대…가계부 작성부터 세금 계산까지”, 안소영 기고, 2017.02.17.

조선일보, “새로운 시장 잡아라! ‘마이데이터’ 쟁탈전”, 김희권 기고, 2020.12.06

중앙일보, “요즘 돈관리 이렇게 하더라…90년생 은행원 폰엔 이것 있다”, 정용환 기고, 2019.12.03.

**손 애 린 (Son, Aelin)**



연세대학교 정보대학원에서 비즈니스빅데이터 트랙 석사과정에 재학 중이며, 주요 연구 분야는 Data Analytics, Data Mining, Digital Business 등이다.

**신 왕 수 (Shin, Wangsoo)**



연세대학교 정보대학원에서 맞춤형 융합 트랙 석사과정에 재학 중이며, 주요 연구 분야는 Data Analytics, Text Mining, Digital Business 등이다.

**이 준 기 (Lee, Zoonky)**



University of Nebraska, Business Administration 에서 근무한 후, 현재 연세대학교 정보대학원 교수로 재직 중이며, 주요 연구 분야는 E-Business Strategy, Big Data Analytics 등이다.

<Abstract>

## **An Analysis of Key Elements for FinTech Companies Based on Text Mining: From the User's Review**

Son, Aelin · Shin, Wangsoo · Lee, Zoonky

### **Purpose**

Domestic asset management fintech companies are expected to grow by leaps and bounds along with the implementation of the “Data bills.” Contrary to the market fever, however, academic research is insufficient. Therefore, we want to analyze user reviews of asset management fintech companies that are expected to grow significantly in the future to derive strengths and complementary points of services that have been provided, and analyze key elements of asset management fintech companies.

### **Design/methodology/approach**

To analyze large amounts of review text data, this study applied text mining techniques. Bank Salad and Toss, domestic asset management application services, were selected for the study. To get the data, app reviews were crawled in the online app store and preprocessed using natural language processing techniques. Topic Modeling and Aspect-Sentiment Analysis were used as analysis methods.

### **Findings**

According to the analysis results, this study was able to derive the elements that asset management fintech companies should have. As a result of Topic Modeling, 7 topics were derived from Bank Salad and Toss respectively. As a result, topics related to function and usage and topics on stability and marketing were extracted. Sentiment Analysis showed that users responded positively to function-related topics, but negatively to usage-related topics and stability topics. Through this, we were able to extract the key elements needed for asset management fintech companies.

**Keyword:** Fintech, Bank Salad, Toss, Text Mining, Topic Modeling, Sentiment Analysis, Aspect-Sentiment Analysis

\* 이 논문은 2020년 11월 25일 접수, 2020년 12월 4일 1차 심사, 2020년 12월 14일 2차 심사, 2020년 12월 22일 게재 확정되었습니다.