

텍스트 마이닝을 활용한 시중 은행과 인터넷 전문은행에 대한 인식 및 기회분석

:시중은행 및 인터넷 전문은행에 대한 리뷰를 바탕으로

윤고은*, 김동해*, 김서영*, 조성윤*

Awareness and opportunity analysis of commercial banks and internet-only banks using text mining

:Based on reviews of commercial banks and internet-only banks

Go-eun Yoon*, Dong-hae Kim*, Seo-young Kim*, Seong-yoon Cho*

디지털 변환의 가속화로 금융 서비스 분야 또한 새로운 변화를 겪고 있다. 이로 인해 인터넷 전문은행 및 디지털 뱅킹에 대한 인식 및 사용현황은 예전과 다를 것으로 예상된다. 시중 은행 및 인터넷 전문은행에 대해 어떻게 인식하고 있는지를 알아보기 위하여 이들의 중요 기능들에 대한 중요도와 만족도를 웹스크래핑을 통해 수집한 텍스트 데이터로 도출한 뒤 최종적으로 기회점수로 수치화하여 도출하는 것이 목표이다. LDA를 통해 은행들의 중요 기능들과 그 기능들의 중요도를 언급량으로 도출하였다. 각 토픽(기능)에 속한 문장들은 자체 제작한 감성 사전을 통해 감성분석을 하여 만족도를 도출했다. 이러한 중요도와 만족도 값을 토대로 기회점수를 도출하여 인터넷 전문은행, 시중은행에 대한 인식과 관련된 인사이트를 얻었다. 이와 같은 연구과정은 인터넷 전문은행과 시중은행 각각이 직면한 문제와 잠재적인 기회에 대한 경영 전략을 제시할 것이다.

Abstract

The acceleration of digital transformation has brought about significant changes in the field of financial services. As a result, perceptions and usage patterns of internet banks and digital banking are expected to differ from the past. To understand how commercial banks and internet banks are perceived, the goal is to extract and quantify the importance and satisfaction of their core functions using text data collected through web scraping. Through the use of LDA, the importance of these functions for different banks was determined based on word frequencies. Sentiment analysis using a custom made sentiment dictionary was employed to assess satisfaction levels of sentences within each topic. Using these importance and satisfaction values, opportunity scores were calculated to gain insights into how internet banks and commercial banks are perceived. This research process aims to provide strategic insights for both internetbanks and commercial banks to address challenges and leverage potential opportunities.

Keywords

crawling, text mining, LDA, topic modeling, segmentation, business strategy

I. 서론

현대 기업 경영에서는 고객의 경영참여가 늘어나며 고객이 경영에서 가지는 영향력이 늘어났다. 최근의 기업 경영에서는 고객의 충족되지 않은 욕구가 무엇인지 찾아 감동을 실현시키는, 진정한 고객 만족이 더욱 강조된다.

은행업을 전자금융거래 방법으로 영위하는 인터넷 전문 은행은 낮은 금리와 편리한 접근성을 바탕으로 시장에서의 점유율을 빠르게 상승시켰다. 더 나아가 코로나 시대의 비대면 업무는 인터넷 전문 은행의 사용률 증가에 큰 영향을 미쳤다. 이러한 흐름에 맞추어 시중은행도 인터넷 뱅킹 등 은행의 디지털 서비스를 강화하며 고객 욕구 충족 및 경쟁력 강화에 전력을 기울이고 있다.

그러나 은행 업무의 어떠한 기능이 고객들의 욕구 충족을 실현시켰는지, 개선해야 하는 분야는 무엇인지 고객 입장에서 수치적인 우선순위를 기업에게 제공하는 연구는 이루어지지 않고 있다.

따라서, 본 연구는 특정 제품의 고객 세분화 방법론을 은행업 전체에 적용하여 시중은행과, 인터넷 전문은행의 기능별 중요도, 만족도에 따라 기회점수를 도출하고자 한다.

II. 데이터 수집 및 전처리

본 연구에서는 시중은행, 인터넷 전문은행 이용자들이 주로 사용하는 fm korea, 네이트판, 레몬테라스, 아프니가 사장이다, 지식인, 행복채테크, 뽐뿌 7개의 커뮤니티 및 카페를 수집 대상으로 선정했다. 데이터 수집 기간은 2019년 1월 1일부터 2023년 7월 31일까지 총 1685일이다. 시중은행, 인터넷 전문은행을 나타내는 키워드를 선정한 뒤 각 사이트의 검색창에 해당 키워드를 검색하여 제목, 본문, 댓글, 날짜를 수집하였다 표1.

표 1.

컬럼명	뜻
Date	리뷰가 쓰인 날짜
Keyword	데이터 수집될 때 사용된 키워드
Title	게시글 제목
Main	게시글 본문
Comment	게시글 본문

수집된 데이터는 인터넷 전문은행 368,231행, 시중은행 444,090행으로 총 812,321행이다. 각 키워드는 ‘토스+뱅크’, ‘토스+주식’과 같이 복합 키워드로 구성했다. 같은 사람이 작성한 리뷰를 구분하기 위하여 ‘Title’열과 ‘Main’열을 하나의 열로 합쳐줬고,

‘Comment’열은 문장별 분리를 수행한 뒤 같은 아이디를 가진 데이터끼리 같은 행으로 합쳐주었다. 이러한 과정을 거쳐 3개의 열을

‘Review’라는 열로 병합하였다. 이에 따라 데이터의 양은 인터넷 전문은행 524,692행, 시중은행 708,913행으로 늘어났다.

데이터 품질 향상을 위해 각 사이트의 크롤링 과정에서 추출되는 불필요한 문자인 HTML 태그, 이모티콘 및 특수 문자를 제거하였다. 데이터 정제를 위해 공백 및 중복 데이터를 제거하였으며 광고성 글을 식별하고 삭제하였다. 광고성 글은 ‘광고’, ‘협찬’, ‘채험단’, ‘무료채험’, ‘채험후기’, ‘예약링크’, ‘신문기사링크’와 같은 키워드가 포함된 데이터를 의미한다.

데이터를 분석 가능한 형태로 가공하기 위해 Mecab 형태소 분석기를 사용하였다. 1000개의 무작위 데이터를 konlpy에서 제공하는 Hannanum, KKma, Mecab, Okt Komoran 으로 각각 형태소 분석을 하여 결과값을 확인 하였을때 Mecab이 속도도 가장 빠르고 토큰화의 품질도 좋아 이를 사용했다.

다음으로 불용어를 제거하였다. 불용어는 텍스트 데이터에서 큰 의미를 가지지 않는 단어로 텍스트 분석의 정확도를 높이기 위해 삭제하였다. 한글에서는 ‘그’, ‘그런’, ‘아’, ‘어’, ‘저’, ‘는’, ‘이’, ‘가’, ‘을’ 같은 의미를 가지지 않는 단어들이 불용어로 분류된다. 이러한 데이터 전처리 과정을 거친 후 시중은행과 인터넷 전문은행 데이터를 별도로 저장하였다. 이에 따라 데이터의 양은 인터넷 전문은행 463,452행, 시중은행 525,160행으로 줄어들었다.

마지막으로 불필요한 데이터를 최대한 제거시키고 은행의 기능을 나타내는 글만 분석에 사용하기 위하여 준지도 학습을 진행하였다. 전체 데이터의 3%인 13,903행, 15,754행을 사용가능한 행과 불가능한 행으로 직접 라벨링을 해준 뒤 DNN(Deep Neural Network) 모델을 사용하여 학습시켜 주었다. 이로써 최종 분석에 사용할 데이터의 양은 인터넷 전문은행425,936행, 시중은행338,287행으로 결정되었다.

III. 데이터 분석 및 결과

본 연구의 목적인 시중 은행과 인터넷 전문 은행의 기능별 기회점수를 도출하기 위해 시중은행 데이터와 인터넷 전문은행 데이터를 여러 세그먼트로 나눈 뒤 세그먼트에 해당하는 데이터를 감성분석 해주었다. 기능의 기회 점수는 각 고객 그룹이 선호하는 제품 기능의 중요도와 만족도를 기반으로 산출된다. 각 고객집단의 중요도와 만족도 점수는 최소-최대 스케일링에 의해 0~10의 척도로 변환된다 (정준각, 해리슨 김, 2023). 분석 과정은 그림 1와 같이 진행된다.

데이터를 여러 세그먼트로 나누는 과정에서 AP-클러스터와 LDA(Latent Dirichlet Allocation) 두 방법이 고려되었지만, AP-클러스터는 인간이 확인하였을 때 같은 주제에 대한 이야기가 아님에도 같은 군집을 형성하는 경우가 많아 최종적으로 LDA를 은행의

기능별로 세그먼테이션을 진행하는 최종 모델로 선정하였다.

그림 1. 기회 점수 도출 프로세스

Fig. 1. Opportunity Scoring Process:



토픽 모델링을 통해 세그먼테이션을 하기 위해 Python의 머신러닝 패키지인 사이킷런(Scikit-Learn)을 활용하였으며 토픽 모델링을 수행하는데 있어서 가장 중요한 하이퍼파라미터인 토픽의 개수를 선정하였다. 토픽의 개수를 선정하기 위해 텍스트 마이닝의 언어 모델을 평가하기 위한 내부 평가 지표 중 하나인 coherence값을 활용했다. 이에 따라 시중은행은 7개의 토픽, 인터넷 전문은행은 6개의 토픽이 최적이라고 판단했다. 시중은행의 6번 토픽은 이벤트, 배송, 배달, 광고 등이 많이 언급된 토픽이고 은행의 기능이 아닌 소상공인들의 개인 사업 고충에 관련된 이야기이다. 시중은행의 7번 토픽은 특정 지을 수 없고, 은행에 관련된 토픽이 아니었다. 인터넷은행의 6번 토픽은 은행이 아닌 다른 여러 제품들에 대한 후기에 관련된 토픽이었다. 이에 따라 시중은행의 6,7번 토픽 인터넷 전문은행의 6번 토픽을 제외하고 각각 5개씩 은행의 기능을 정의하였다.

표 1. 시중은행 LDA 각 토픽별 군집 결과

Table 1. LDA Clustering Result for Commercial Bank

토픽	기능	사용자 정의	포함 단어
토픽1	주요 서비스	은행의 주요 서비스인 카드, 계좌, 결제, 통장, 체크카드 등에 대한 사용 경험을 나누는 일반 사용자 모임	카드 계좌 결제 통장 발급
토픽2	투자 서비스	은행 서비스를 투자에 적극적으로 활용하고 정부의 반응을 관찰하는 사용자 모임	상품 지급 정부 투자 정책
토픽3	일반 및 자산 대출 서비스	은행 서비스 중 대출을 부동산 자산 투자에 적극적으로 활용하며 신용,금리에 민감한 사용자 모임	대출 신용 금리 연체 전세
토픽4	기업 대출 서비스	은행 서비스 중 대출을 기업단위로 사용하는 사업자 사용자 모임.	신청 대출 보증 기업 차
토픽5	은행 점포 서비스	시중은행의 실물 점포, 지점을 적극적으로 활용하는 사용자 모임	전화 지점 직원 센터 직접

표 2. 인터넷 전문은행 LDA 각 토픽별 군집 결과

Table 2. LDA Clustering Result for Internet Bank

토픽	토픽	사용자 정의	포함 단어
토픽1	사업자 대출 서비스	은행의 대출 서비스를 적극적으로 활용하는 젊은 사업자 또는 개인 사업자 모임.	신청 전화 보증 서류 사장
토픽2	일반 대출 서비스	은행의 대출 서비스를 전세, 생활비 등에 활용하는 일반 사용자 모임	대출 신용 상품 금리 전세
토픽3	투자 서비스	은행 서비스를 주식, 사업 등 자산 증대에 활용하는 사용자 모임	투자 주식 사업 수익
토픽4	디지털 बैं킹 및 결제 서비스	은행의 결제 서비스를 적극적으로 활용하며 디지털 बैं킹에 관련된 서비스도 활용하는 사용자 모임	계좌 결제 통장 현금 입금
토픽5	모바일 결제 서비스	모바일 결제와 포인트 서비스를 활용하는 사용자 모임. 모바일 앱을 설치하고 사용하며 관리 및 서비스 제공에 대한 정보를 찾고자 함	페이 결제 할인 무료 업체

LDA를 한 결과값으로 중요도를 도출하였다. 사용자들의 각 토픽에 대한 언급 빈도수를 계산한 뒤 정규화 시킨 후 시그모이드 함수를 적용시킨 다음 10을 곱하여 중요도 점수를 도출하였다.

다음으로 각 기능에 따른 감성분석을 진행해주었다. 감성사전은 직접 제작한 감성 사전을 사용하였다. 감성 사전에 들어있는 토큰(단어)별 감성 점수의 범위는 -2부터 +2 까지이다. 기능에 따른 감성점수는 각 토픽별로 데이터를 나눈 뒤 불용어의 지정을 처음한 전처리와 다르게 한 불용어 사전을 활용하여 토큰화 한 다음 문장별로 감성점수를 계산하였다. 토큰(단어)별 점수를 문장 내의 위치에 따라 가중치를 달리하여 계산했다. 그 후 단어별 감성점수의 평균을 문장의 감성점으로 도출하였다. 문장 별 감성점수는 토픽별로 평균을 구한 뒤 시그모이드 함수를 사용하여 0~1 사이 값으로 만든다음 10을 곱하여 만족도 점수를 도출하였다.

마지막으로 $opp = imp_score_i + \max(imp_score_i - sent_score_i, 0)$ 공식에 따라서 기회 점수를 계산하였다. 기회점수의 범위는 0부터 20까지이다(정준각, 해리슨 김, 2023).

표 3. 시중은행 기회점수

Table 3. Commercial Bank Opportunity Score

기능	중요도	만족도	기회 점수
주요 서비스	5.2785	7.6470	5.2785
투자 서비스	5.4036	4.1528	6.6544
일반 및 자산 대출 서비스	5.3216	6.6568	5.3216
기업 대출 서 비스	5.3653	4.8590	5.8716
은행 점포 서 비스	5.3743	6.6404	5.3743

표 4. 인터넷전문은행

Table 4. Internet Bank Opportunity Score

기능	중요도	만족도	기회 점수
사업자 대출 서비스	5.4061	6.0866	5.4061
일반 대출 서 비스	5.6545	6.4213	5.6545
투자 서비스	5.4044	5.8033	5.4044
디지털 뱅킹 및 결제 서비 스	5.3318	6.1759	5.3318
모바일 결제 서비스	5.3589	6.3993	5.3589

IV. 결론 및 시사점

1. 결론

시중은행은 앱 사용성과, 고객 편의성 향상을 통한 인식 개선이 고객 욕구 충족을 위해 우선적으로 필요하다. 시중은행이 높은 만족도를 보일 것으로 예상한 기업대출 서비스는 불편한 접근성과 복잡한 대출 승인 서비스로 인해 더 낮은 점수가 도출되었다. 가장 높은 기회점수가 도출된 투자서비스 또한 비슷한 맥락이었다. 높은 기회점수의 요인으로 느리고 불안정한 어플리케이션과 가시성이 떨어지는 UI등이 있었다. 인터넷 전문은행의 경우, 모든 기능들에 대한 고객 및 사용자들의 만족도가 중요도 보다 높은 값을 보여준다. 이는 고객들이 인터넷 전문은행인 카카오뱅크, k뱅크, 토스뱅크의 주요 기능들에 모두 만족한다는 뜻이고 이는 인터넷 전문 은행의 사용률 증가에 큰 영향을 미쳤을 것이다. 이러한 높은 만족도의 가장 중요한 요인으로 은행의 기능간의 유기적인 연동성이 있었다. 이는 디지털 전환에 박차를 가하고 있는 시중은행이 위의 분석 결과에 따라 기회점수가 높은 값부터 해결하고 결국에는 각 기능간의 유기적인 연동성을 높일 수 있는 방안을 다방면에서 노력해야 할 것이다.

2. 한계점 및 시사점

1) 한계점

본 연구의 한계점은 은행 기능에 기회점수를 도출할 때 은행 기능별 중요도에 대한 계산을 단순히 언급량으로 했다는 점이다. 은행 기능에 대한 별점이 포함된 리뷰가 있는 데이터가 및 소스가 있었다면 보다 더 정확한 값을 제공할 수 있었을 것이다.

2) 시사점

본 연구는 현재 국내의 시중은행과 인터넷전문은행의 경영 전략 및 기능 및 서비스 개선 방안을 고객의 참여가 증가하는 현대 경영에 맞추어 고객 중심적 접근을 통한 분석을 했다는 것에 있다.

추가적으로 어떠한 특정 제품의 리뷰데이터로 고객 세분화를 할 때 사용하던 방법론을 참고하여 기업, 더 나아가 시장 자체에 적용하여 유의미한 기회 점수를 도출해서 개선점을 제시했다는 점에서 큰 의의가 있다.

참고문헌

- [1] 정준각, 해리슨 김(2023). 온라인 제품 리뷰를 통한 신제품 개발을 위한 고객 세분화를 위한 해석 가능한 머신러닝 기반 접근 방식. *International Journal of Information Management*. 70 호.
- [2] 김동규, 이동욱, 박장원, 오성우, 권성준, 이인용, 최동원(2022 June28). KB-BERT 금융 특화 한국어 사전학습 언어모델과 그 응용. *Journal of Intelligent Information Systems*. 191~206
- [3] 고객의 투자상품 선호도를 활용한 금융상품 추천시스템 개발_김효중, 김성범, 김희웅
- [4] 홍길동 (2014). 한국 언론의 새로운 지평 연구. <한국언론연구>, 35권 1호, 23-54.