Investigating the Factors Influencing User Churning Behavior in Spam Filtering Apps: A Comparison between Churners and Users and Big Data Analysis of App Logs
스팸 필터링 앱 사용자의 이탈 요인 연구: 이탈자와 지속 사용자 비교 및 앱 로그 빅데이터 분석을 중심으로

Ae Ri Lee(First Author)
Dept. of Business Administration, Sangmyung University
(*sharon@smu.ac.kr*)
Chanhee Kwak(Corresponding Author)
Dept. of Industrial Data Science, Kangnam University
(*chk@kangnam.ac.kr*)

Recently, various spam calls and messages targeting mobile phone users have been continuously increasing. As the types of spam are becoming diversified with increasing frequency, it not only causes considerable inconvenience to mobile users, but also becomes a risk factor in information security and financial accidents. Telecommunication companies have developed and distributed spam filtering apps, and mobile users are increasingly interested in spam blocking and spam information sharing services. Nevertheless, there have been behaviors that do not use or stop using spam filtering apps. Therefore, it is important for companies and organizations, which need to increase the usage rate of spam filtering apps, to diagnose the use status of spam filtering apps and to analyze how customer churn occurs. This study compares the usage patterns of spam filtering apps between users and churners, and identifies the characteristic behaviors of churners. In particular, by analyzing the big data of app usage log, we derive the factors that affect the increase in churn rate. With the results of this study, it contributes to establishing a safer mobile phone environment by preventing customers from churning and activating the use of spam filtering functions.

Key Words: Mobile Spam Filtering App, Big Data Analysis, Customer Churn Factor

Submission Date: 07. 27. 2020 Revised Date: (1st: 10. 07. 2020) Accepted Date: 10. 21. 2020

Ⅰ. 서 론

정보통신기술(ICT) 발전과 함께 모바일 네트워킹 환경이 보편화 되면서, 모바일 폰 사용자를 타깃팅 하는 각종 스팸(spam) 전화 및 메시지가 지속적으로 증가하고 있다(DigitalTimes, 2019; Roy et al., 2020). 방송통신위원회 통계자료에 의하면, 2016년 하반기 휴대전화 스팸 신고 건은 약 1,307만 건이었고, 2019년 8월에는 약 2,433만 건으로 1.9배 증가한 것으로 나타났다(Korea Communication Commission, 2019). 최근 사용자가 원치 않게 받게 되는 대출 권유 및 보험 가입 유도, 상품판매, 주식/투자 홍보 등의 광고성 스팸의 종류는 점차 다양해지고 그 빈도 또한 증가하고 있어 모바일 서비스 이용자들에게 상당한 불편을 초래하고 있다(Park and Park, 2013; Roy et al., 2020). 또한, 불법의 영역에 들어가는 도박, 불법의약품, 심지어는 불법대출 및 사금융 까지 모바일을 광고 수단으로 이용하면서(Korea Communication Commission, 2019), 이용자들의 불편을 넘어 불안까지 일으킬 수 있다(Osho et al., 2014). 더구나 스팸 메시지나 스팸 전화가 범죄 형태로 진화된 스미싱 및 피싱 사고까지 발생하고 있어, 개인의 정보 보안 및 금융 측면에서의 위험 요인이 되고 있다(Almeida et al., 2011; Hong, 2019).

이와 같이 모바일 사용자를 대상으로 증가하는 스팸을 막기 위해, 한국인터넷진흥원(KISA) 등 국가기관에서는 스팸대응센터를 설치하고, 통신사들은 실시간 스팸 필터링 앱(app)을 개발 및 보급해 왔으며, 최근 모바일 사용자들의 스팸 차단 및 스팸 정보 공유 서비스에 대한 관심이 더욱 커지고 있다(Jiang et al., 2016; Jung et al., 2019). 스팸 필터링 앱은 단순히 스팸성 광고를 막는 것뿐만 아니라 이용자들의 편익을 높이고, 불법적인 해킹 위험으로부터의 안전을 도모한다. 현재 한국의 모바일 스팸 필터링 앱은 크게 후후(whowho)와 T전화로 양분되는 형태를 띠고 있으며, 사용자들은 두 가지의 앱을 동시에 쓰기도 한다.

이러한 스팸 필터링 앱에 대한 니즈가 증가함에도 불구하고, 아직까지 스팸 필터링 앱을 사용하지 않거나 사용을 중단하고 이탈하는 현상이 나타나고 있다. 그러므로 스팸 필터링 앱 이용률을 높이고자 하는 기업 및 기관 입장에서는 사용자들의 스팸 필터링 앱 이용 현황을 파악하고, 이탈 현상이 어떠한 패턴과 원인으로 발생하는지 분석하는 것이 매우 중요할 것이다. 한편, 최근 객관적 데이터 분석에 기반한 의사결정의 중요성이 부각되면서, 기업들은 데이터 경영 및 비즈니스 인텔리전스(business intelligence) 환경 구축을 위해 체계적인 빅데이터 분석 기법을 도입하고자 노력하고 있다(Jang, 2015; Kim and Kim, 2019).

본 연구는 스팸 필터링 앱 이용과 관련한 빅데이터를 분석하여 모바일 스팸 필터링 앱 사용의 이탈 현상을 진단하고자 한다. 이를 통해 밝히고자 하는 연구 질문은 다음과 같다.

- 스팸 필터링 앱 이용 행태 데이터 분석을 통해, 스팸 필터링 앱을 지속적으로 사용하는 이용자와 이탈자 그룹 간 서로 다른 이용 패턴을 발견할 수 있는가?
- 이탈률을 높이는 주요한 영향 변수는 무엇인가?

본 논문에서는 빅데이터 분석을 통해, 위의 연구 질문에 대한 답을 제시함으로써 스팸 필터링 앱 이탈을 방지하고 사용률을 높이기 위한 시사점을 제공

하고자 한다.

Ⅱ. 기존문헌연구

현재까지 수행된 스팸 필터링 앱과 관련된 연구들은 제한적이다. 관련 선행연구들을 살펴보면 다음과 같다.

2.1 스팸 필터링을 위한 기술적 대응 방안 연구

스팸 필터링 앱 및 시스템과 관련한 선행연구 중 대다수는 스팸 필터링 알고리즘에 대한 연구들이다.

Ahn et al.(2015)은 POI(Point of Interest) 검색 알고리즘을 활용하여 문자 메시지 안의 스팸 문자열을 찾고 스팸 메시지 여부를 구별하는 안티스팸(anti-spam) 앱을 설계하고 구현함으로써, 모바일 사용자들이 스팸 판단 결과와 메시지 내용을 확인하고 스팸 메시지일 경우 삭제할 수 있는 방안을 제안하였다. Byun and Kim(2014)은 메시지 다이제스트 알고리즘을 이용하여 자동으로 스팸 데이터베이스를 구축하는 스팸 메시지 필터링 시스템을 연구하였다. Lee et al.(2014)은 스미싱 차단을 위해 안드로이드 플랫폼 기반 스마트폰에서 스미싱 관련 악성 앱이 문자 메시지를 탈취하기 전 메시지를 수집할 수 있는 모듈을 추가하고, 사용자 기기 데이터베이스(database) 안에 수집한 수신 문자 메시지가 정상적으로 저장되었는지를 확인하여 악성 앱이 문자 메시지를 탈취하여 서버로 전송하는 것을 차단하는 모듈을 안드로이드 커널에 삽입하는 방법을 제안하였다.

최근 기계학습(machine learning) 및 딥러닝(deep learning) 기술 발전에 따라, 이를 스팸 필터링 알고리즘에 적용한 연구들이 등장하고 있다. Joe and Shim(2009)은 서포트 벡터 머신(Support Vector Machine, SVM) 기반의 스팸 필터링 시스템을 제안하였는데, SVM 분류기로의 학습을 통해 문자 메시지의 스팸 인식률을 높이고자 하였다. 이들 연구자들은 휴대폰 사용자들의 스팸 또는 비 스팸 문자 메시지를 수집하여 학습 데이터와 테스트 데이터를 구축하고, 학습 데이터로부터 특징 벡터를 추출한 후 이들 특징을 이용하여 만들어진 SVM 분류기를 통해 학습하도록 하였다. 단, 제안된 시스템은 PC 환경에서 구현되었다. Lee and Kang(2018)은 자연어 처리를 위한 텍스트 분석 기법인 워드 임베딩 기법과 딥러닝 기법을 이용하여 스팸과 정상 문자 메시지로 분류하는 방법을 제안하고, 기존의 문자 메시지 필터링 방법과 성능을 비교 연구하였다. Roy et al.(2020)은 문자 메시지 안의 스팸 필터링을 위해 컨볼루션 신경망(Convolutional Neural Network, CNN)과 장단기 메모리 방식(Long Short-Term Memory, LSTM) 모델 기반의 딥러닝 기법을 활용하여 기존 방식 대비 정확도를 높이고자 하였다. Lee and Choi(2011)는 사용자의 휴대 단말에서 수신한 메시지 데이터에 대한 기계학습 시, 나이브 베이즈(Naïve Bayes)와 선형(linear) SVM 알고리즘을 적용하고, 스마트폰에서 자원 효율적으로 동작할 수 있는 스팸 필터링 분류기를 개발하여 제시하였다. Sethi et al.(2017)은 스팸 문자 메시지 탐색에 활용될 수 있는 다양한 기계학습 알고리즘을 비교 분석하였고, 오픈 공공 데이터 세트를 이용하여 이들 기계학습 알고리즘들이 서로 다르게 스팸 메시지를 분류하고 이에 따른 정확도의 차이가 있음을 설명하였다.

이들 선행연구들은 주로 스팸 문자 메시지를 판별

하고 차단할 수 있는 필터링 알고리즘을 제안함으로써 스팸으로부터 사용자의 불편을 줄이고 스미싱 사고를 예방할 수 있는 기술적 개발 방안을 제시하고자 하였다.

2.2 스팸 필터링 서비스 관련 사용 행태 및 사례 조사 연구

일부 선행연구에서는 스팸 필터링 서비스와 관련한 사용 행태 분석과 사례 조사를 수행하였다.

Jung et al.(2019)은 스팸 차단/공유 서비스의 유형을 크게 3가지 형태, 즉, OEM 연동 모드, 팝업 모드, 전체 화면 모드로 분류하고, 서비스의 사용자 경험(UX)에 대한 디자인 측면의 만족도 점검과 개선점을 제안하였다. 연구자들은 팝업 모드일 경우는 스마트폰의 기본 화면을 가리는 사용 상의 불편함이 있고, 전체 화면 제공 방식에서는 다양한 테마 등 독자적인 서비스 제공에는 유리하나 스마트폰 단말 UX와 이질적인 서비스 UX가 제공된다는 한계점을 지적하였다. 결과적으로, 단말 기본 화면 상에서 제공되는 링크방식의 OEM 연동 모드가 단말 화면과 일관된 UX를 제공함으로써 사용자 만족도를 높일 수 있다고 설명하였다.

Park and Park(2013)은 스마트폰에서 스미싱 해킹 공격으로 인해 비정상적인 인터넷 결제가 발생하는 피해 사례가 증가하고 있음을 지적하고, 스미싱 해킹 공격의 원리와 사례 분석을 하여 스미싱을 이용한 안전결제 시스템 피해 예방에 대한 보안 방안을 제시하였다. 보안 방안으로는, 고객센터에서 불법적인 URL을 신고받아 스팸 필터링 시스템에서 접속 차단, 동일 IP에서 복수 소액 결제 발생 시 해당 IP 결제 차단, 스미싱 피해 발생 시 결제 취소, 1회 결제 한도나 월간 결제 한도 축소 방안 등이 제시되었다.

Wang et al.(2010)은 스팸 메시지가 모바일 폰 사용자의 일상을 심각하게 방해하고 통신사에게 악영향 끼치고 있다고 지적하면서, 사용 행태 기반의 소셜 네트워크 분석(behavior-based social network analysis)과 시간적 스팩트럴 분석(temporal and spectral analysis) 기법을 결합하여 스패머(spammer)와 합법적인 메시지 발신자를 구별하도록 하는 단문 메시지(SMS) 안티-스팸 시스템의 개선 방안을 제안하였다.

Bin et al.(2016)은 스팸 메시지 탐지가 오늘날 모바일 통신 사업자들의 큰 과제이며, 실무적으로 스패머 탐지 구현과 실 적용에의 어려움과 이슈가 있음을 지적하였다. 그 이슈의 예로는, 스패머를 포함한 사용자들의 개인정보보호 및 프라이버시 이슈로 인해 네트워크 사업자가 전체 규모의 콘텐츠 기반 SMS 스팸 탐지 기술(whole-volume content based SMS spam detection techniques)을 쉽사리 사용할 수 없다는 것과 낮은 정밀도로 인해 자동 필터링 후 수동적인 검토를 많이 수행해야 한다는 것이다. 연구자들은 이러한 이슈를 개선하고 보완하기 위하여, 행태 분석(behavior analysis)을 기반으로 한 SMS 스패머 탐지 방안을 제안하였다. 이 연구에서는 모바일 메시지 사용 행태 특성 데이터(calling frequency, average calling interval, top 1/2/3 called No. frequency ratio, called No. number 등)를 기계학습 훈련 데이터에 반영하여 스패머 탐지 정확도를 높일 수 있는 방안이 제시되었다.

Yeon et al.(2019)은 스마트폰 출시 시 선탑재 되는 앱의 효과와 선탑재 앱들 중 고객들이 선호하는 앱이 무엇인지 조사하고, 사용자들이 몇 개 정도의 선탑재 앱을 원하는지 등의 행태 정보를 분석함으로써

사용자 니즈에 맞춘 스마트폰 선탑재 앱의 구성 방안을 제안하였다. 해당 연구 결과, 앱 사용량 측면에서 앱의 선탑재 효과는 상당히 큰 것으로 조사되었으며, 여러 종류의 단말 사용자들이 공통으로 선호하는 선탑재 앱에는 고객센터, 멤버십, 보인 인증 앱과 더불어 스팸 전화/문자 필터링 앱이 포함되었다. 또한 선탑재를 원하는 앱의 평균 개수는 8개 이하였으며, 비록 대중적으로 인기 있는 앱이라도 선탑재를 원하는 비율이 낮게 나타났다. 이를 통해, 꼭 필요한 필수 앱으로만 구성하고, 선탑재 앱 개수를 최소화하고자 하는 사용자 니즈가 있음을 보여주었다.

Kim et al.(2019)은 새로운 정보통신기술이 가져온 이점을 좀 더 효율적으로 사용하고 실용성을 높이기 위해 미디어 리터러시 또는 디지털 리터러시의 중요함을 강조하고, 모바일 환경에서 갖추어야 할 미디어 리터러시의 구성 요소가 무엇인지 조사하였다. 조사 결과를 기반으로 모바일에서의 미디어 리터러시 측정 모델을 제시하고, 모바일 미디어 리터러시에 있어 세대 간 격차가 있음을 검증하였다. 이 연구를 통해 정리된 모바일 미디어 리터러시 중, 모바일 이용 능력 측정 지표에는 앱 설치 능력, 스마트기기 환경 설정 역량, 무선 네트워크 설정 등이 포함되어 있으며, 차단 기술 능력 지표에는 스팸 메일 또는 스팸 문자 차단 능력, 유해 콘텐츠 차단 앱 사용 가능 역량, 광고성 팝업 차단 능력 등이 포함되어 있다. 연구 결과, 스팸 및 유해 콘텐츠 차단 앱 사용 역량을 비롯하여 모바일 미디어 리터러시의 거의 모든 영역에서 세대 간 격차가 나타났다.

Hong(2019)은 세계 각국에서 발행하고 있는 피싱 범죄에 대응하는 국가별 대응 정책을 비교 연구하였다. 조사 결과, 최근 국가별로 피싱에 대한 정부 차원의 관심과 대응 수준이 높아지고 있으며, 특히 예방 차원에서의 교육과 홍보 등을 통해 적극적인 선제적 대응을 1단계로 하여, 2단계로는 신속한 피싱 신고 접수와 실시간 대응이 강조되고 있다. 예를 들어 대만의 경우는, 통화 도중 피싱이 의심될 때 버튼만 누르면 경찰이 실시간 통화 내용을 감청하고 조치가 가능하도록 하였고, 사용자들에게 실시간 주의 메시지를 보내거나 음성 경고 시스템을 보다 개선하는 등 강력한 대응 조치를 취하고 있다. 또한 2단계에서 그치는 것이 아니라 이후 3단계에서는 피싱 범죄의 재범 방지, 피해자 보호 및 피해 회복을 위한 관련 기관의 노력이 증가하고 있다고 설명하였다.

상기에서 살펴본 선행연구들의 주요 연구 내용 및 결과를 한눈에 파악하기 쉽도록 다음 〈Table 1〉로 요약 정리하였다. 이들 기존연구들은 스팸 필터링 서비스 및 차단 앱의 중요성과 개선점을 다양한 관점(UX 및 앱 선호도 관점, 스패머 식별 측면에서의 행태 분석, 그룹/국가 간 비교 등)에서 조사 분석하였다. 단, 기존 연구에서는 스팸 필터링 서비스 사용 활성화를 위한 사용자 측면에서의 이용 행태 분석과 실제 로그 데이터 기반의 이탈 요인 연구가 이뤄지지 않았다. 또한, 아직까지 스팸 및 피싱 방지를 위한 서비스와 앱 이용 관련 연구 자체가 매우 부족한 실정이다.

따라서 본 연구에서는 스팸 필터링 앱 사용 현황을 실 로그 데이터 분석을 통해 진단하고, 스팸 필터링 앱을 지속적으로 사용하는 이용자와 이탈자 그룹 간 사용 패턴의 차이가 있는지 분석하고자 한다. 특히 이탈 확률을 높이는 주요 영향 요인을 파악함으로써 스팸 필터링 앱 사용률을 높일 수 있는 방안을 제시하고자 한다.

〈Table 1〉 스팸 필터링 서비스 사용 행태 및 사례 분석 관련 기존문헌 정리

저자	연구 내용	스팸 필터링 관련 주요 연구 결과
Jung et al. (2019)	스팸 차단/공유 서비스의 화면 UX 유형에 따른 사용자 만족도 차이 분석	스팸 차단/공유 서비스 제공 시, 단말 화면과 일관된 UX를 제공함으로써 사용자 만족도 향상 가능함
Park and Park (2013)	스미싱 해킹 공격 사례 분석 및 안전 결제시스템 피해 예방 방안 제시	스미싱으로 인한 비정상적 결제 피해를 방지할 수 있도록 스팸 필터링 시스템 등을 통한 실질적인 접속 차단 및 결제 제한선 설정 정책 등이 제시됨
Wang et al. (2010)	사용 행태 기반 소셜 네트워크 분석 기법 활용한 안티-스팸 시스템 개선 방안 제시	사용 행태 기반의 소셜 네트워크 분석과 시간적 스팩트럴 분석 기법을 통합적으로 활용함으로써 스팸 메시지 식별 수준을 높일 수 있음
Bin et al. (2016)	사용 행태 분석 기반 SMS 스패머 탐지 개선 방안 연구	모바일 메시지 사용 행태 특성 데이터를 활용한 기계학습을 통해 스패머 탐지 정확도 향상 가능함
Yeon et al. (2019)	스마트폰 출시 시 선탑재 앱의 효과와 고객 선호 선탑재 앱 조사	폰에 앱 선탑재 시, 사용량 증가 효과가 큼; 특히 스팸 전화/문자 필터링 앱은 고객 선호도가 높아 선탑재가 필요한 것으로 조사됨
Kim et al. (2019)	모바일 미디어 리터러시 측정 모델 제시 및 세대 간 격차 분석	모바일 미디어 리터러시 측정에 있어, 스팸 메일/문자 차단과 유해 콘텐츠 차단 앱 사용 역량이 주요한 지표가 됨; 이들 차단 앱 사용에 있어 세대 간 격차가 나타남
Hong (2019)	국가별 피싱 범죄 발생 현황 및 대응 방안 조사	세계적으로 피싱 범죄 방지 및 대응 정책이 강화되고 있으며, 1단계(예방 교육 및 홍보)에서 2단계(신속 신고접수 및 실시간 대응), 3단계(재범방지, 피해자 보호/회복 조치)까지 피싱 대응 수준이 높아지고 있음

Ⅲ. 연구방법론

3.1 데이터 수집

본 연구에서는 모바일 스팸 필터링 앱 사용 로그 데이터를 활용하여 분석을 진행하였다. 스팸 필터링 앱 제공사인 A기업의 협조 하에, 2019년 5월 7일부터 5월 20일까지의 데이터를 확보하였다. 이 기간 동안 총 4,825,907개의 데이터가 수집되었다. 수집된 데이터 중 지속 이용자 데이터는 4,584,396개 이고, 이탈자 데이터는 241,511개 이다. 이 중 이상치(outlier)와 불완전 데이터를 제외하여 데이터를 정제하였고, 지속 이용자 4,473,691개 (95%), 이탈자 233,408개 (5%)의 데이터를 분석에 사용하였다. 수집된 데이터에서, 지속 이용자(다수 클래스)의 수에 비해 이탈자(소수 클래스)의 수가 월등히 적다. 이러한 불균형 데이터(imbalanced data)의 문제를 해결할 수 있는 다양한 방법을 문헌에서 확인할 수 있는데, 크게 Undersampling과 Oversampling으로 나눌 수 있다. Undersampling의 경우 다수 클래스에서 일정 부분을 랜덤하게 추출(Random Undersampling) 하거나 클래스의 경계를 활용 하는 등의 기법(TomekLink)으로 다수 클래스의 수를 소수 클래스와 동일하게 맞춘다(Batista et al., 2004; Prusa et al., 2015). 반면, Oversampling은 소수 클래스를 랜덤하게 중복적으로 추출(Random Oversampling) 하거나, 대표 클래스별로 중복 추출

(Stratified Oversampling), 혹은 소수 클래스 데이터를 활용하여 인공적인 소수 클래스를 만드는 기법(GAN, SMOTE) 등이 있다(Farquad et al., 2012; Oh et al., 2019). 두 가지 중 어떤 샘플링 기법이 다른 것보다 우월하다고 할 수 없으며, 각 데이터의 특성에 맞는 샘플링 기법을 사용해야 한다.

본 연구에서는 Random Undersampling 기법을 사용하였는데, 소수 클래스의 수가 200,000이 넘는 충분한 수이고, 이 수와 동일하게 다수 클래스를 랜덤하게 추출하더라도 통계적 타당성을 잃지 않기 때문이다. 또한, 이탈자의 행태를 탐구하고자 하는 본 연구의 목적을 고려했을 때, 이탈자의 정보를 최대한 활용하는 것이 적합하다고 판단하였다. 단, 본 연구에서는 다수 클래스의 특징을 Undersampling 된 데이터에 반영하기 위해 층화(Stratified) 기법을 사용하였다. 본 연구의 데이터 중 층화에 사용할 수 있는 최신 OS 여부, 최신 앱 버전 여부의 두 항목을 고려하여 층화 추출을 진행하였다. 최신 OS는 안드로이드 기준 9.0 버전 이상을 기준으로 하였고, 최신 앱 버전은 앱 스토어에서 다운로드 받을 수 있는 최신 버전의 앱을 기준으로 하였다.

본 연구에서 활용된 분석용 데이터 구조는 아래 〈Table 2〉와 같다.

〈Table 2〉 분석용 데이터 구조 정의

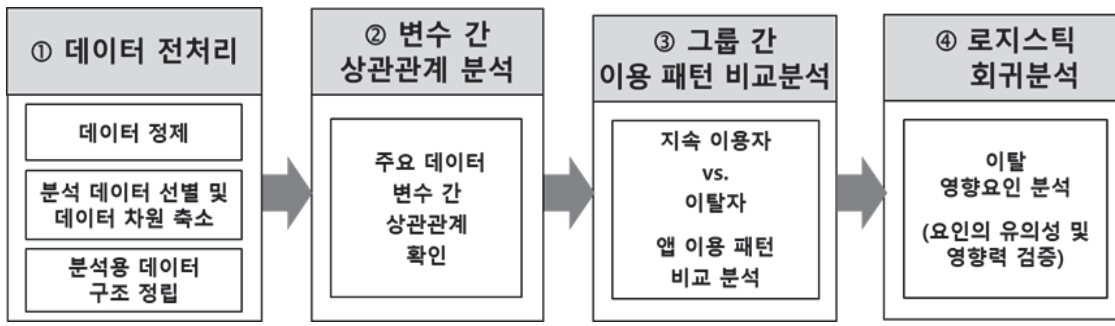
구분	필드명	설명
식별자	Id	데이터 구분을 위한 식별 번호
이탈 여부	Is_leave	1: 이탈, 0: 지속 사용
날짜 변수	Date	데이터 생성(측정)일
	LastDatetime	가장 최근 API 호출일
	PutDatetime	스팸 필터링 앱 서비스 가입일
사용자 기기/앱 정보 관련 변수	AppVersion	앱 버전
	OSType	스마트폰 운영체제(OS) 종류
	OSVersion	스마트폰 운영체제(OS) 버전
	PhoneModel	스마트폰 모델
스팸 정보 공유/등록 관련 변수	SPP (Spam Phone Number Putting)	스팸 번호 등록 (횟수)
	RPP (Reliable Phone Number Putting)	안심 번호 등록 (횟수)
	ISP (Information Sharing Putting)	정보 공유 등록 (횟수)
	SBS (Spam Blocking Sharing)	스팸 차단 공유 (횟수)
문자/전화 사용 관련 변수	SMR (Spam Message Receiving)	스팸 문자 수신 (횟수)
	NMR (Normal Message Receiving)	일반 문자 수신 (횟수)
	NCS (Normal Call Sending)	일반 전화 발신 (횟수)
	SCS (Spam Call Sending)	스팸 전화 발신 (횟수)
	NCR (Normal Call Receiving)	일반 전화 수신 (횟수)
	SCR (Spam Call Receiving)	스팸 전화 수신 (횟수)

〈Table 2〉 분석용 데이터 구조 정의 (계속)

구분	필드명	설명
스팸 필터링 앱 사용을 위한 주요 옵션 설정 관련 변수	ConfigHPositive (Configuration with Highly Positive Option)	스팸 차단 관련 중요도가 높은 옵션 기능을 사용하도록 설정 변경 정도 • 스팸번호 모두 차단 옵션 ON • 스마트 자동 스팸 차단 옵션 ON • 실시간 스팸 차단 옵션 ON • 알림창 사용 및 기본 앱으로 설정
	ConfigHNegative (Configuration with Highly Negative Option)	스팸 차단 관련 중요도가 높은 옵션 기능을 사용 안하도록 설정 변경 정도 • 스팸번호 모두 차단 옵션 OFF • 스마트 자동 스팸 차단 옵션 OFF • 실시간 스팸 차단 옵션 OFF • 기본 앱으로 설정 안함
	ConfigMPositive (Configuration with Moderately Positive Option)	스팸 차단 관련 중요도가 중간 정도 되는 옵션 기능들을 사용하도록 설정 변경 정도 • 전화 수신 시, 모든 번호 (또는 모르는 번호)에 대한 스팸 차단/확인 기능 사용 • 전화 발신 시, 모든 번호 (또는 모르는 번호)에 대한 스팸 차단/확인 기능 사용 • 통화 종료 시, 모든 번호 (또는 모르는 번호)에 대한 스팸 차단/확인 기능 사용 • 부재중 시, 모든 번호 (또는 모르는 번호)에 대한 스팸 차단/확인 기능 사용 • 문자 알림창 실행 시, 모든 번호 (또는 모르는 번호)에 대한 스팸 차단/확인 기능 사용 • 단말 OEM 연동 중 전화 알림 통계전송 설정
	ConfigMNegative (Configuration with Moderately Negative Option)	스팸 차단 관련 중요도가 중간 정도 되는 옵션 기능들을 실행 안하도록 설정 변경 정도 • 전화 수신 시, 모든 번호에 대한 스팸 차단/확인 기능 실행 안함 • 전화 발신 시, 모든 번호에 대한 스팸 차단/확인 기능 실행 안함 • 통화 종료 시, 스팸 확인 기능 실행 안함 • 부재중 시, 스팸 차단/확인 기능 실행 안함 • 문자 알림창 실행 시, 모든 번호에 대한 스팸 차단/확인 기능 실행 안함 • 전화 알림 통계전송 미설정

3.2 연구 설계 및 분석 절차

데이터 분석 절차는 다음 〈Figure 1〉과 같이 진행되었다. 사용 로그 데이터를 바탕으로 1) 데이터 전처리, 2) 변수 간 상관관계 분석, 3) 지속 이용자 및 이탈자 그룹 간 이용 패턴 비교 분석, 4) 로지스틱 회귀분석(이탈확률을 높일 수 있는 영향 변수 분석) 순서로 진행하였다.



〈Figure 1〉 데이터 분석 절차

1) 데이터 전처리 단계에서 우선 부적절한 오류 데이터를 정제 처리하였다. 부적절한 오류 데이터 예로는, 데이터 생성일(Date)이 스팸 필터링 앱 서비스 가입일(PutDatetime) 보다 더 빠른 날짜로 표기되어 있는 경우, 가장 최근 API 호출일(LastDatetime)이 2199/12/31로 표시되어 있는 경우, 앱 버전(AppVersion)이 None인 경우 등이며, 이들 데이터가 분석용 데이터에 포함되지 않도록 삭제하였다. 데이터 정제 후, 본격적인 분석 데이터로 사용될 변수를 선택하도록 하였다.

데이터 분석에 사용될 변수의 수가 너무 많아지게 되면 차원의 저주(curse of dimensionality)에 빠질 수 있는데, 이는 너무 많은 변수로 인해 데이터의 차원이 희소해지게 되고, 모델의 과적합으로 이어질 수 있다. 따라서 본 연구에서는 차원의 저수에 빠지지 않도록 모든 로그 데이터를 다 분석에 사용하지 않고, 분석에 꼭 필요한 주요 데이터 변수를 선별하였다. 본 연구의 목적은 스팸 필터링 앱 지속 이용자와 이탈자(즉, 앱을 더 이상 사용하지 않는 사람) 간의 앱 이용 패턴의 차이점이 있는지 분석하고 이탈 확률을 높일 수 있는 변수가 무엇인지 파악하는 것이므로, 이와 관련된 로그 데이터 변수가 선택될 수 있도록 하였다. 본 연구의 데이터 수집은 스팸 필터링 앱 서비스 제공 회사와의 협력 하에 이뤄졌으며, 이 과정에서 스팸 필터링 서비스에 대한 기업의 사업지식(business domain knowledge)을 활용하여 분석이 필요한 주요 데이터 변수를 비즈니스 전문가와 함께 선별하였고, 기존에 분리되어 있었던 일부 변수들을 통합하는 등 데이터 전처리 작업을 하였다. 예를 들어, Config(앱 사용 옵션 설정) 관련 변수 설명 안에 있는 세부적인 옵션들이 각각 다른 데이터 변수들로 분리되어 있었는데, 전처리 과정을 통해 이를 통합된 변수로 조정하였다. 특히 '앱 사용 옵션 설정'에 있어 스팸 차단 관련 중요도가 높은지 또는 중간 수준인지와 긍정적(우호적)인지 또는 부정적인 것인지 대해서 A사 담당자와의 논의를 통해 결정하였다. 이러한 처리 과정을 통해 꼭 필요한 특징을 포함하도록 데이터의 차원을 축소(dimensionality reduction)하는 등 사전 조정을 한 후 분석이 이뤄질 수 있도록 하였다. 또한, 일부 극단적인 사용자들로 인한 추정치 왜곡 가능성을 최대한 줄이고자 사용 변수들에 대해 Box-Cox Transformation을 시행하였고, 이를 통해 이상치(outlier)들의 효과를 줄임과 동시에 분석 시행 시 더 안정적인 결과를 도출할 수 있도록 하였다.

또한 Config 관련 변수를 중요도를 기준으로 재통

합하였다. ConfigHPositive와 ConfigHNegative의 차이를 ConfigH로, ConfigMPositive와 ConfigMNegative의 차이를 ConfigM으로 계산하여 분석(상관관계분석 및 회귀분석)에 활용되도록 하였다. 이는 각 옵션의 설정이 이진(binary)구조로 되어 있어 ON과 OFF를 동시에 설정할 수 없기 때문이며, 데이터의 차원을 줄이는 효과도 기대할 수 있다.

이와 같은 데이터 전처리 단계를 거쳐 분석용 데이터가 준비되도록 하였다.

2) 상관관계분석 단계에서는 연구의 변수 간 상관관계를 사전에 확인하도록 하였다. 아래 〈Table 3〉은 본 연구에 사용된 변수들의 상관관계를 나타내고 있다. 상관관계분석 결과, 이후 분석에 영향을 줄 만한 높은 상관관계를 보이는 경우는 없었다. 이에, 다중공선성(multicollinearity) 이슈가 없는 것을 확인하였다.

3) 그룹 간 비교 분석 단계에서는 지속 이용자와 이탈자의 스팸 필터링 앱 사용을 위한 주요 옵션 설정 관련한 이용 패턴을 비교 분석하였다. 이러한 비교 분석을 통해, 이탈자들이 스팸 필터링 앱 사용 시 보이는 행태 및 이탈 확률이 높은 사전 시그널이 무엇인지 파악하도록 하였다. 이에 대한 상세 분석 결과는 4.1에서 설명하도록 한다.

4) 마지막 단계에서는 로지스틱 회귀분석을 실시하였다. 로지스틱 회귀분석을 통해, 종속변수(결과변수)인 이탈 여부에 유의한 영향을 주는 독립변수(원인변수)를 분석하고, 특히 이탈에의 영향력이 높은 변수가 무엇인지 파악하도록 하였다. 이에 대한 상세 분석 결과는 4.2에서 설명하도록 한다.

Ⅳ. 분석 결과

4.1 그룹 간 이용 패턴 비교 분석 결과

스팸 필터링 앱 지속 이용자와 이탈자 이용 패턴

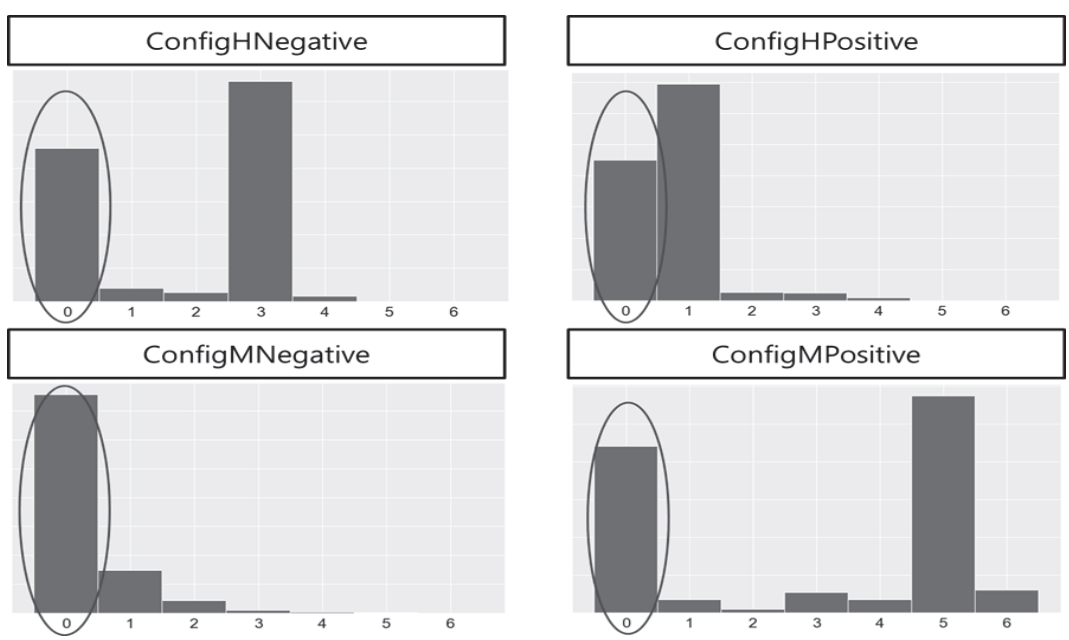
〈Table 3〉 변수 상관관계

	1	2	3	4	5	6	7	8	9	10	11
1.SBS											
2.SPP	0.497										
3.RPP	0.015	0.013									
4.ISP	0.091	0.183	0.033								
5.NMR	0.041	0.009	0.011	0.003							
6.SMR	0.185	0.050	0.011	0.010	0.264						
7.NCS	-0.028	0.014	0.023	0.003	0.495	0.196					
8.SCS	0.016	0.012	0.013	0.005	0.204	0.202	0.213				
9.NCR	0.046	0.029	0.013	0.005	0.338	0.180	0.293	0.131			
10.SCR	0.065	0.043	0.010	0.004	0.299	0.247	0.227	0.175	0.482		
11.CfgM	0.017	0.014	-0.003	-0.001	0.015	0.052	0.077	0.039	0.215	0.193	
12.CfgH	0.011	0.036	0.013	0.012	-0.049	-0.018	0.002	-0.019	-0.106	-0.108	-0.456

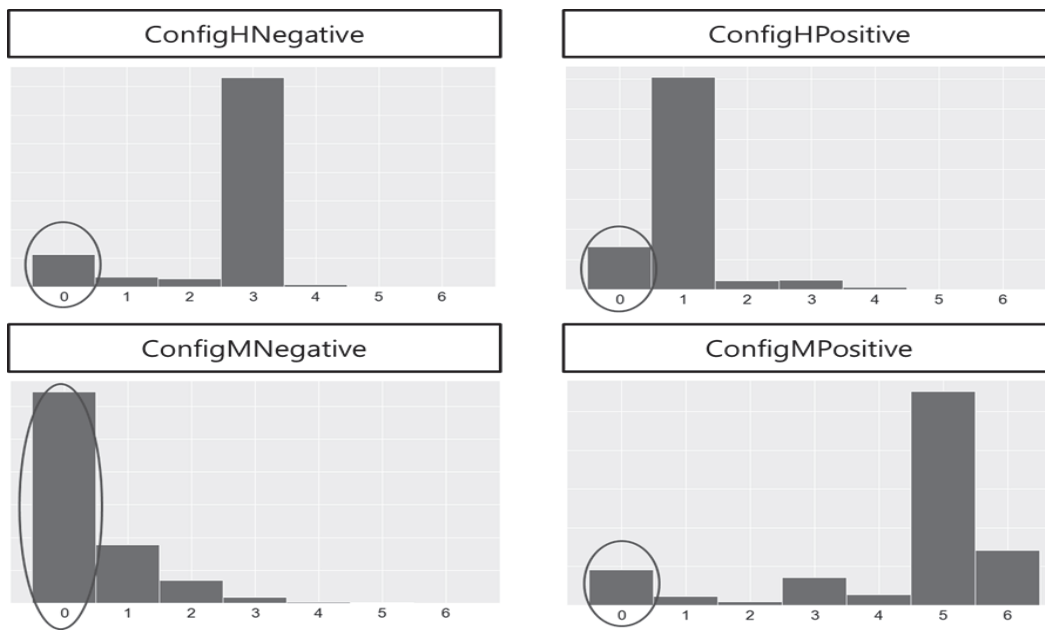
을 비교하기 위해, 몇 가지 변수들을 기준으로 그룹 간 비교 분석을 수행하였다.

〈Figure 2〉와 〈Figure 3〉은 이탈자와 지속 이용자의 옵션 설정을 비교한 결과이다. 이탈자의 경우는 전체적으로 스팸 필터링 앱 사용을 위한 주요 옵션 설정 변경 값들이 0인 비율이 높게 나타났다. 반면, 지속 이용자의 경우는 전체적으로 주요 옵션 설정 변경 값들이 0인 비율이 이탈자 대비 낮게 나타났다. 〈Figure 2〉와 〈Figure 3〉의 주요 옵션 설정 변경 값 분포 패턴을 비교해 보면, 이탈자의 경우는 ConfigHNegative, ConfigHPositive, ConfigMNegative, ConfigMPositive 변수 모두에서 설정 변경 값이 0인 비율이 지속 이용자 대비 특히 높게 나타난 것을 알 수 있다. 비록 4개 변수의 설정 변경 값 빈도가 서로 유사한 것도 있지만(예: ConfigHPositive의 값이 1인 빈도가 두 그룹 모두 많음), 가장 두드러진 특징은 이탈자의 경우는 이들 주요 옵션 설정 변경 값이 0인 경우가 많다는 것이다. 이는 이탈자의 경우, 스팸 필터링 앱 사용에 큰 관심이 없어 주요 옵션 설정 값에 대한 변경 행위 자체를 잘 안하는 것으로 파악된다. 반면, 지속 이용자의 경우는 자신의 앱 서비스 사용 경험의 최적화를 위해 이들 주요 옵션 설정 값을 변경해 가면서 사용한다고 볼 수 있다.

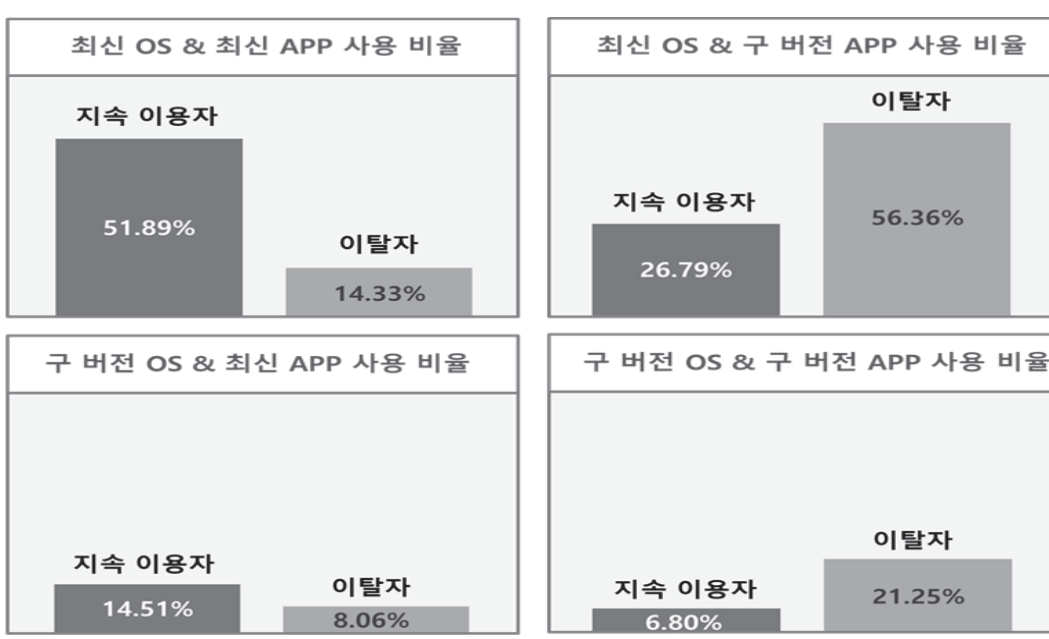
다음으로, 최신 OS와 최신 앱(APP)을 기준으로 지속 이용자와 이탈자 그룹을 비교 분석 하였다. 그 결과는 〈Figure 4〉와 같다. 먼저, 최신 OS와 최신 앱을 사용하는 지속 이용자의 수가 가장 많은 것을 알 수 있다. 전체 지속 이용자의 51.89%가 최신 앱과 최신 OS를 사용하는 것으로 나타난 반면, 같은 상황에서는 이탈자는 14.33%의 수치를 보였다. 반면, 구 버전 OS와 구 버전 앱을 사용하거나(이탈자



〈Figure 2〉 이탈자 그룹의 주요 옵션 설정 변경 값 분포



〈Figure 3〉 지속 이용자 그룹의 주요 옵션 설정 변경 값 분포



〈Figure 4〉 OS 및 APP 최신 버전 사용 비율 비교

의 21.25%), 최신 OS와 구 버전 앱을 사용하는 경우(이탈자의 56.36%)는 이탈자의 비율이 상대적으로 높은 것을 알 수 있다. 즉, 사용 중인 OS 버전과 앱 버전은 이용자의 이탈률 예측에 중요한 시그널 요인이 될 수 있다.

4.2 로지스틱 회귀분석(Logistic Regression) 결과

본 연구에서는 Python의 Logit 함수를 이용하여, 로지스틱 회귀분석을 실시하였다. 로지스틱 회귀분석은 이변량의 값을 갖는 종속변수와 독립변수들 간의 인과관계를 추정하기 위한 분석 기법이다(Hosmer and Lemeshow, 1989). 본 연구에서는 종속변수(결과변수)를 이탈 여부(1: 이탈, 0: 지속 사용)로 설정하고, 로지스틱 회귀분석을 통해 이탈 발생 확률을 높이는 주요한 요인 변수가 무엇인지 예측할 수 있도록 하였고, 분석 결과는 〈Table 4〉에 정리되어 있다. 〈Table 4〉에서 Model I은 기본적인 로지스틱 회귀분석 모델이며, Model II는 규제항을 적용한 분석 결과이다. 규제항을 적용함으로써, 데이터 분석에서 나타날 수 있는 과적합을 줄이고, 모델의 일반적인 추론 능력을 향상시킬 수 있다(Ng, 2004). 특히 L1 규제는 여러 특성 변수 중 해당 모델에서 중요한 특성 변수가 무엇이고 그 영향력 및 효과가 어느 정도인지 설명하기 쉽도록 한다(Ng, 2004). 본 연구에서는 L1 규제항(L1 Regularization)을 적용하여 분석하였다. 또한 모든 분석에서 Robust Error를 사용함으로써, 분석의 정당성을 확보하고자 노력하였다(Whilte, 1980). 분석 결과, SBS(스팸 차단 공유)를 제외한 모든 변수가 통계적으로 유

〈Table 4〉 로지스틱 회귀분석 결과

	Model I (BC)		Model II (BC-Regularized)	
독립 변수	회귀계수	p-value	회귀계수	p-value
SBS	0.0231n.s.	0.5390	0.0231n.s.	0.5390
SPP	0.5294***	0.0000	0.5294***	0.0000
RPP	3.6982***	0.0000	3.6983***	0.0000
ISP	1.8410***	0.0000	1.8419***	0.0000
NMR	-0.3792***	0.0000	-0.3792***	0.0000
SMR	0.3720***	0.0000	0.3720***	0.0000
NCS	0.0348***	0.0000	0.0348***	0.0000
SCS	0.8676***	0.0000	0.8676***	0.0000
NCR	0.0832***	0.0000	0.0832***	0.0000
SCR	-0.2769***	0.0000	-0.2769***	0.0000
ConfigM	0.0878***	0.0000	0.0878***	0.0000
ConfigH	-0.0167***	0.0000	-0.0167***	0.0000
Robust Error	Yes			

***: $p < 0.001$, n.s.: not significant, 종속변수: Is_leave (이탈 여부), BC: Box-Cox Transformation

의한 것으로 나타났다.

회귀분석 결과를 통해 요인들이 이탈에 미치는 영향력 정도를 파악할 수 있다. 먼저, 부(-)의 회귀계수 값을 보이는 경우가 있었는데, 이들 요인들은 이탈이 아닌 지속 사용 발생 가능성을 높이는 변수로 해석될 수 있다. 여기에 해당되는 변수는 NMR(일반 문자 수신), SCR(스팸 전화 수신), ConfigH(스팸 차단 관련 중요도 높은 옵션 사용 설정 정도) 이다.

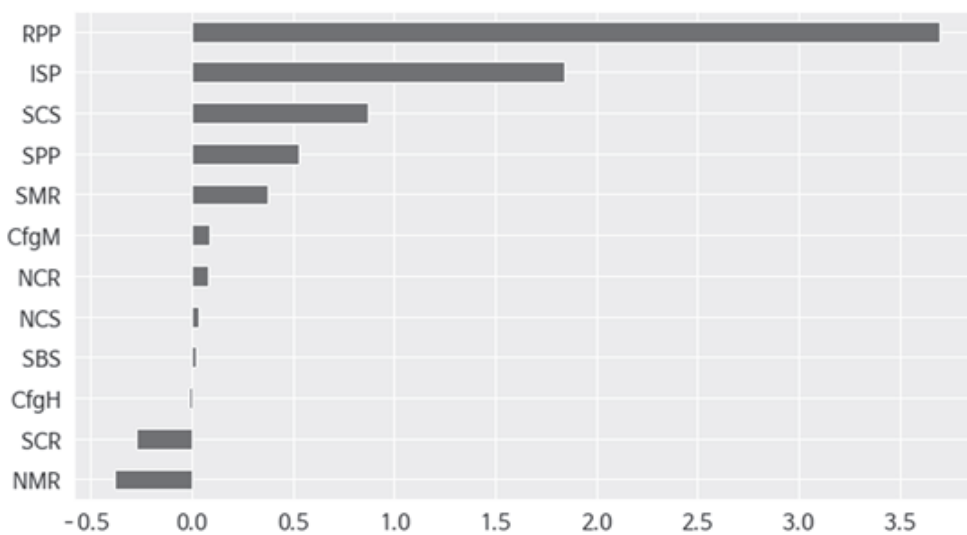
반면, 정(+)의 회귀계수 값이 클수록 이탈 확률을 높이는 주요한 영향 변수로 볼 수 있는데, 정의 계수 값이 큰 순서는 RPP(안심 번호 등록), ISP(정보 공유 등록), SCS(스팸 전화 발신), SPP(스팸 번호 등록), SMR(스팸 문자 수신), ConfigM(스팸 차단 관련 중요도가 중간 수준의 옵션 사용 설정 정도), NCR(일반 전화 수신), NCS(일반 전화 발신) 순이다.

분석된 회귀계수 값을 시각적으로 비교하기 용이하도록 아래 〈Figure 5〉로 정리하였다.

Ⅴ. 결과 토의 및 시사점

본 연구는 스팸 필터링 앱 지속 이용자와 이탈자의 사용 패턴을 비교 분석하여 이탈자들이 스팸 필터링 앱 사용 시 보이는 특징적인 이용 행태를 파악하고, 이탈률을 높이는 주요 요인을 규명하는 것을 목적으로 하였다. 스팸 메시지 및 스팸 전화는 모바일 폰 사용자들의 일상을 방해할 뿐만 아니라 보안 및 금융 사고의 위험을 유발하기에, 스팸 차단의 중요성이 날로 증가하고 있지만 이와 관련된 연구들은 아직 부족한 실정이다. 이러한 상황 가운데, 본 논문은 실제 이용 행태 로그 빅데이터를 분석하여 이탈과 관련된 요인들을 탐색한 연구로서, 저자들이 알기로는 아직까지 이와 같은 연구가 수행된 바 없었다.

따라서 본 연구는 스팸 필터링 앱 사용 이탈과 관련된 빅데이터 분석을 시도한 초기 연구로서 학술적 의의가 있다. 또한 기존의 설문조사 및 실험 연구가



〈Figure 5〉 회귀계수 값 비교

아닌, 실 데이터를 기반으로 분석한 연구로서, 현실을 보다 정확하게 진단하고 연구의 실효성을 높였다고 할 수 있다.

본 연구에서 그룹 간 비교 분석 결과, 이탈자는 지속 이용자 대비 전체적으로 스팸 필터링 앱 사용을 위한 주요 설정 값 변경을 하지 않는 경우가 많음을 알 수 있었다. 이를 통한 시사점은, 주요 옵션 기능 설정 값을 이용하는 정도를 보면 이탈 가능성이 높은 사람들이 누구인지 알 수 있다는 것이다. 즉, 주요 설정 값 변경 행위 정도가 매우 낮은 사람들을 대상으로 스팸 필터링 앱 사용의 중요성에 대한 인식을 제고하고 이탈을 방어하는 전략이 필요하다. 한편, 분석 결과를 통해 스팸 필터링 앱 지속 이용자의 경우는 자신의 사용 경험 최적화를 위해 앱 설정 값을 변경하면서 사용하고 있음을 알고 있고, 심지어 ConfigHNegative(스팸 차단 관련 중요도 높은 옵션을 사용 안하도록 설정 변경) 값이 큰 경우에도 이후 지속적으로 앱을 이용하는 것으로 파악되었다. 단, 지속 이용자 중 ConfigHNegative 값이 큰 경우는 앱 사용에 있어 부정적인 경험을 하고 설정을 변경한 경우에 해당될 수 있으므로 이에 대한 대책 마련이 필요하다고 할 수 있다. 사용 OS 및 앱 버전에 대한 그룹 비교 결과, 최신 앱을 설치하였을 때 이탈 확률이 매우 줄어드는 것을 확인할 수 있었다. 지속 이용자들은 최신 OS와 최신 앱을 사용하는 비율이 가장 높았으며, 이탈자의 경우는 최신 OS를 사용함에도 불구하고 구 버전의 앱을 사용하는 비율이 가장 높았다. 특히 이탈자 그룹의 경우, 구 버전 앱 사용 비율이 약 78%에 달했다. 이와 같이 구 버전의 앱을 그대로 사용하는 경우 스팸 필터링 앱에 대한 관심도가 낮아 결국 이탈로 이어질 가능성이 존재한다. 따라서 이들을 대상으로 앱 업그레이드(upgrade)를 유도하는 이벤트 등을 추진함으로써 지속 이용률을 높일 수 있도록 노력해야 할 것이다.

로지스틱 회귀분석 결과, 이탈 확률을 높이는 변수와 낮추는 변수가 구별되어 나타났다.

먼저, 이탈 확률을 높이는 변수 중에는 RPP(안심번호 등록), ISP(정보 공유 등록), SPP(스팸 번호 등록)가 있었다. 사용자가 안심하다고 생각되는 번호를 앱에 직접 등록해야 하거나 정보를 공유하도록 등록하고 스팸 번호를 손수 등록한다는 것은, 스팸 필터링 앱의 자동적인 필터링 기능이 잘 작동되지 않아 스팸이 아닌 번호가 스팸으로 잘못 차단되거나 제대로 스팸 필터링이 되지 않는 경우가 발생할 수 있다는 것이다. 물론 현재 100% 완벽하게 스팸을 판별하여 차단할 수 있는 앱 기능을 제공하는데 한계가 있기에, 이러한 사용자의 직접 등록/공유 기능을 지원하는 것은 필요한 조치이다. 그러나, RPP, ISP, SPP 값이 큰 경우는 사용자가 이러한 등록 행위를 하기 전/후에 스팸 필터링 앱 사용에 대한 좋지 않은 경험(즉, 제대로 필터링이 안되는 경험)을 했을 가능성이 존재하므로 이에 대한 면밀한 검토와 보완이 필요한 것으로 판단된다. 그리고 본인의 전화로 스팸 전화를 발신하는 SCS 값이 높은 사용자는 스패머일 가능성이 있으므로 스팸 필터링 앱 사용을 기피할 수 있다.

분석 결과, 이탈률을 높이는 변수에 SMR(스팸 문자 수신)이 있었다. SMR 값이 큰 경우는 스팸에 해당되는 문자를 많이 받은 경우이다. 단, SMR 값이 큰 경우에는, "사용자가 모든 문자를 일단 수신하겠다고 옵션을 설정했을 때, 스팸 가능성이 높은 번호로부터 수신된 문자임을 표시해 주면서 문자가 수신되는 경우" 또한 포함될 수 있다. 그러므로 관련 앱 개발사는 스팸 차단 기능 구현 시 어떠한 형태로 UX를 설계하고 구현하는 것이 사용자의 불편을 줄이고 이탈을 방지할 수 있는지 세심하게 검토해야 할 것이다.

특히, ConfigM(스팸 차단 관련 중요도 중간 수준의 옵션 사용 설정 정도)이 이탈 확률을 높이는 변수로 나타났는데, 이는 원래 기대했던 바와 다른 결과이다. 중요도 중간 정도의 옵션 기능에 해당되는 경우는 전화 수/발신, 통화 종료, 부재중, 문자 알림창 실행 시 모든 번호에 대한 스팸 차단 및 확인 기능을 사용토록 설정하는 것과 통계정보를 전송토록 설정하는 부분이다. 이러한 옵션은 스팸 필터링 앱의 정교화 및 지속적인 서비스 통계 분석을 위해서 필요한 기능이나, 사용자 측면에서는 부담이 될 수도 있다. 본 연구 결과는, 스팸 필터링 앱이 제공하는 옵션 기능 중 사용자 단말에 부하를 유발하거나 사용자를 오히려 불편하게 만드는 과도한 기능이 있는지 점검할 필요가 있음을 시사한다. 따라서 스팸 필터링 앱 개발사에서는 이러한 점을 고려하여 옵션 기능들을 재검토해 볼 필요가 있다.

한편, 분석 결과, 이탈 확률을 낮추는 변수에는 NMR(일반 문자 수신) 및 ConfigH(스팸 차단 관련 중요도 높은 옵션 사용 설정 정도)와 함께 SCR(스팸 전화 수신)이 포함되었다.

SCR 값이 큰 경우는, 전화 수신 시 스팸이라고 표시되는 것이 많은 경우가 이에 해당된다. 즉, SCR 값이 클수록 이탈률이 적을 수 있다는 것은, 그 만큼 이 스팸 필터링 앱이 스팸 전화를 잘 필터링 한다는 것을 사용자가 눈으로 실감함으로써 앱의 유용성을 인식하고 지속 사용하게 될 수 있다고 해석될 수 있다.

연구 결과, ConfigH 값이 양(+)의 값으로 클수록 이탈 확률이 낮아지는 것을 확인하였다. 이를 반대로 이야기하면, ConfigH 값이 음(-)인 경우, 즉, ConfigHNegative 값이 큰 사람은 앱 사용 이탈 가능성이 높은 사용자이므로 이들에 대한 이탈 방어 정책을 구사해야 할 것이다.

본 연구에는 몇 가지 한계점이 있다. 첫째, 본 연구에서는 약 14일간의 로그 데이터를 확보하여 분석하였는데, 향후 연구에서는 로그 데이터 수집 기간을 보다 확대하여 분석함으로써 이탈자의 행태적 특징을 보다 면밀히 파악할 수 있을 것이다. 특히, 로그 데이터 기반 고객 이탈 요인 분석에 대한 최신 연구에서는 행동 데이터를 시계열적으로 분석하여 새로운 시사점을 제시하고 있다. 따라서 데이터 수집기간을 확대한 향후 연구에서는 시계열 분석을 통해 고객 변화를 새롭게 고찰할 수 있을 것이다. 둘째, 본 연구는 실제 이용 데이터를 분석한 측면에서 장점이 있지만, 빅데이터 분석의 특성 상 이탈자의 내면적인 심리적 요인까지 파악할 수 없는 한계가 있다. 향후 연구에서는 이들 이탈자에 대한 추가적인 인터뷰 및 설문조사를 수행하여 이를 보완할 수도 있을 것이다.

결론적으로, 본 연구를 통해, 통신사 등에서 이미 선제적으로 제공하는 스팸 필터링 앱 사용 시 이탈 요인들이 무엇인지 파악하고, 이탈 전에 보이는 사전 이용 행태에서의 특징을 살펴볼 수 있었다. 본 연구 결과를 토대로, 스팸 필터링 앱 사용의 이탈을 방지하고 스팸 차단기능 사용을 활성화시킴으로써 보다 안전한 모바일 환경 조성에 기여할 수 있을 것이다.

참고문헌

Ahn, H. Y., Cho, W. Z., and Lee, J. W.(2015), "Implementation of A Mobile Application for Spam SMS Filtering Using Set-Based POI Search Algorithm," *Journal of Digital Contents Society,* 16(5), pp.815-822.

Almeida, T. A., Hidalgo, J. M. G., and Yamakami, A.(2011), "Contributions to the study of

SMS spam filtering: new collection and results," *In Proceedings of the 11th ACM Symposium on Document Engineering*, pp. 259-262.

Batista, G. E., Prati, R. C., and Monard, M. C. (2004), "A study of the behavior of several methods for balancing machine learning training data," *ACM SIGKDD Explorations Newsletter*, 6(1), pp.20-29.

Bin, Z., Gang, Z., Yunbo, F., Xiaolu, Z., Weiqiang, J., Jing, D., and Jiafeng, G.(2016), "Behavior analysis based SMS spammer detection in mobile communication networks," *In 2016 IEEE First International Conference on Data Science in Cyberspace (DSC)*, pp.538-543.

Byun, S. M., and Kim, J.(2014), "Spam Message Filtering System using Message Digest Algorithm," *In Proceedings of KIIT Conference*, pp.120-123.

DigitalTimes, "Over 130 million spam reports in the last 5 years…010 Outgoing rate surge," 2019, Available at http://www.dt.co.kr/contents.html?article_no=2019092902109931032002.

Farquad, M. A. H., and Bose, I.(2012), "Preprocessing unbalanced data using support vector machine," *Decision Support Systems*, 53(1), pp.226-233.

Hong, S. S(2019), "A Comparative Study on the Phishing Fraud Prevention," *The Police Science Journal*, 14(1), pp.101-130.

Hosmer, D., and Lemeshow, S.(1989), *Applied Logistic Regression*, Wiley & Sons, New York.

Jang, Y. J.(2015), "Big Data, Business Analytics, and IoT: The Opportunities and Challenges for Business," *The Journal of Information Systems*, 24(4), pp.139-152.

Jiang, M., Cui, P., and Faloutsos, C.(2016), "Suspicious behavior detection: Current trends and future directions," *IEEE Intelligent Systems*, 31 (1), pp.31-39.

Joe, I. W., and Shim, H. T.(2009), "A SVM-based spam filtering system for short message service (SMS)," *The Journal of Korean Institute of Communications and Information Sciences*, 34(9B), pp.908-913.

Jung, S. K., Seo, M. H. I., Park, I. S., and Park, H. S.(2019), "A Study on User Experience Design of Spam Block/Sharing Service", *In Proceedings of The HCI Society of Korea*, pp.1018-1021.

Kim, K. H., Kim, G. J., and Lee, S. J.(2019), "Media Literacy Components and Generation Gap in the Mobile Environment," *Korean Journal of Broadcasting*, 33(4), pp.55-36.

Kim, S. S., and Kim, Y. J.(2019), "An Empirical Study on Users Intention to Use Insurtech Digital Insurance Platform Service," *Korean Management Review*, 48(4), pp.997-1043.

Korea Communications Commission(2019), *Voice spam, business operator transmission volume increases, user reception decreases*, Korea Communications Commission.

Lee, H. Y., and Kang, S. S.(2018), "SMS text messages filtering using word embedding and deep learning techniques," *Smart Media Journal*, 7(4), pp.24-29.

Lee, S. J., and Choi, D. J.(2011), "Personalized mobile junk message filtering system," *The Journal of the Korea Contents Association*, 11(12), pp.122-135.

Lee, S. Y., Kang, H. S., and Moon, J. S.(2014), "A study on smishing block of android platform environment," *Journal of The Korea Institute of Information Security & Cryptology*, 24

(5), pp.975-985.

Ng, A. Y.(2004), "Feature selection, L1 vs. L2 regularization, and rotational invariance," *In Proceedings of the twenty-first international conference on Machine learning*.

Oh, J. H., Hong, J. Y., and Baek, J. G.(2019), "Oversampling method using outlier detectable generative adversarial network," *Expert Systems with Applications*, 133, pp.1-8.

Osho, O., Ogunleke, O. Y., and Falaye, A. A.(2014), "Frameworks for mitigating identity theft and spamming through bulk messaging," *In 2014 IEEE 6th International Conference on Adaptive Science & Technology (ICAST)*, pp.1-6.

Park, I. W., and Park, D. W.(2013), "A study of intrusion security research and smishing hacking attack on a smartphone," *Journal of the Korea Institute of Information and Communication Engineering*, 17(11), pp. 2588-2594.

Prusa, J., Khoshgoftaar, T. M., Dittman, D. J., and Napolitano, A.(2015), "Using random undersampling to alleviate class imbalance on tweet sentiment data," *In 2015 IEEE International Conference on Information Reuse and Integration*, pp.197-202.

Roy, P. K., Singh, J. P., and Banerjee, S.(2020), "Deep learning to filter SMS spam," *Future Generation Computer Systems*, 102, pp.524-533.

Sethi, P., Bhandari, V., and Kohli, B.(2017), "SMS spam detection and comparison of various machine learning algorithms," *In 2017 International Conference on Computing and Communication Technologies for Smart Nation (IC3TSN)*, pp.28-31.

Wang, C., Zhang, Y., Chen, X., Liu, Z., Shi, L., Chen, G., and Lu, W.(2010), "A behavior-based SMS antispam system," *IBM Journal of Research and Development*, 54(6), pp.1-16.

White, H.(1980), "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity," *Econometrica: Journal of the Econometric Society*, 48(4), pp.817-838.

Yeon, B. H., Kang, W. Y., and Choi, S. J.(2019), "A Study on the Configuration of Pre-install Applications on Smartphone for Customer Needs," *Journal of Broadcast Engineering*, 24(1), pp.105-117.

• The author Ae Ri Lee is an assistant professor in the Department of Business Administration at Sangmyung University. She worked at Korea Telecom as a senior manager in Business Planning and R&D Division. She received her Ph.D. in Information Systems from Yonsei University and her MBA in Technology Management from KAIST. Her research interests include digital transformation, business intelligence, and information security & privacy. She has published papers in Information & Management, Computers in Human Behavior, Internet Research, Behaviour & IT, and Journal of Global Information Management.

• The author Chanhee Kwak is an assistant professor in the Department of Industrial Data Science at Kangnam University. He received Ph.D. degree in management engineering from Korea Advanced Institute of Science and Technology. His research interests include data analytics, privacy and information systems, and digital transformation. His research has been published in International Journal of Information Management and Journal of Knowledge Management.