

# 토픽 모델링과 머신 러닝 방법을 이용한 온라인 C2C 중고거래 시장에서의 사기 탐지 연구

## A Study on the Fraud Detection in an Online Second-hand Market by Using Topic Modeling and Machine Learning

이 동 우 (Dongwoo Lee) 알티데이터랩 연구원  
민 진 영 (Jinyoung Min) 조선대학교 경상대학 경영학부 부교수, 교신저자

### 요 약

온라인 C2C 중고거래에 대한 수요가 증가하고 있으나 물품을 보내지 않거나 명시한 것과 다른 물건을 보내는 방식으로 부당한 금전적 이득을 챙기려는 사기 행위자들의 수도 증가하고 있다. 본 연구는 이러한 사기를 미연에 방지하기 위한 머신 러닝 방법을 이용한 사기 탐지 모델을 구축하였다. 이를 위해 대표적 C2C 중고거래 플랫폼인 중고나라에서 145,536건의 거래 게시글을 수집하였다. 이후 이들 게시글에서 토픽 모델링 기법을 이용하여 상품 설명 내용의 주제를 추출하였으며, 상품 설명의 언어적 특성, 준언어적 특성, 상품의 특성, 게시글의 포스팅 특성, 구매자 특성, 거래 특성들을 추출하였다. 이를 XGBoost 방법에 기반한 머신 러닝 모델을 구축하여 사기 게시글을 탐지하였다. 분석 결과, 사기 게시글은 글 자체의 길이가 대체로 짧고, 제공하는 정보가 적고 상대적으로 구체적이지 않은 것으로 나타났으며 명사를 상대적으로 적게 쓰고 이미지도 사용하지 않거나 적게 사용하는 글이 대부분인 것으로 나타났다. 또한 상대적으로 숫자와 공백의 비율이 높게 나타났으며 정상 게시글의 경우 명사의 경우 상품의 정보, 동사의 경우 전달, 형용사의 경우는 행위와 관련된 단어들이 사용되었으나 사기 게시글은 뚜렷한 주제를 가지지 못하는 것으로 나타났다. 본 연구는 전화번호나 계좌번호를 사용하는 기존의 방법과 달리 다양한 게시글의 특성으로 사기 여부를 탐지하는 모델을 구축했다는 점에서 학술적, 실무적 시사점을 가지고 있다.

**키워드** : C2C 중고거래, 기계학습, 사기탐지 모델

## I. 서 론

전자상거래 형태 중 기업이 아닌 개인이 판매자

와 구매자가 되는 C2C(Consumer to Consumer) 시장은 주로 새 상품 위주로 거래가 되는 B2C(Business to Consumer)나 B2B(Business to Business) 시장과 달리 중고품 거래가 활발하게 일어나는 시장이다. 중고거래란 상품 가치의 일부 또는 전부를 보존하

† 본 논문은 이동우의 2021년도 석사 학위논문을 재구성 및 보강하여 작성되었음.

여 다시 쓸 수 있도록 시장에 내놓아 거래하는 것을 말하는데(Xue and Yang, 2010), 저렴한 가격으로 제품을 구매하려 하거나, 절판되었거나 재고 부족 등의 이유로 상품을 구하기 힘든 상황에서 더 높은 금액을 지불하더라도 구입하려는 경우 등 B2C시장에서 충족되지 못한 소비자들이 C2C 중고거래로 눈길을 돌리고 있다(이보한, 나중연, 2020). 이런 수요에 힘입어 중고거래 시장은 지난 20년간 지속적으로 증가하였고 최근 들어서는 전 세계적으로 폭발적인 성장세를 보이고 있다(Lemel, 2020). 미국의 경우 2016년 기준 2,900만 달러(Fernando *et al.*, 2018), 일본의 경우 1조엔 가량에 달하고 있으며(Yamamoto and Ohshima, 2017), 중국의 경우는 2018년 기준으로 7,000억 위안을 넘어섰다(KOTRA, 2020). 우리나라 중고 시장의 경우 가장 큰 규모의 플랫폼인 ‘중고나라’의 거래액은 2019년 기준 3조 5,000억 원이며, 국내 전체 규모는 10조 원 이상으로 추산되고 있다(조유빈, 2020). 특히, 최근 COVID-19 확산의 영향으로 장기적인 경제적 위기가 예상되고 오프라인 거래를 기피하는 경향이 심화되며, 보다 저렴한 가격으로 비대면 중고 거래를 하려는 수요가 더욱 증가하고 있는 추세이다(김진홍, 안희동, 2020).

그러나 C2C 중고거래는 판매자가 판매하는 제품의 품질을 통제하기 어렵기 때문에 거래 안전에 대한 문제가 발생할 수 있다(이보한, 나중연, 2020). 즉, 판매자가 일방적으로 정보를 제공하는 정보의 비대칭 문제로 인해 구매자는 제품의 품질, 적정 가격 등에 대한 정보를 전적으로 추정해야만 하고 거래의 연속성 또한 없기 때문에 구매자의 피해가 발생하고 있다는 것이다(Dimoka *et al.*, 2012; Li and Chen, 2020). 피해자들이 직접 피해사례를 올려 운영되는 The Cheat라는 사이트에 따르면 대표적인 중고 거래 사이트 3곳인 중고나라, 번개장터, 당근마켓에서 2020년 등록된 피해건수만 16만 건에 달하며(더치트, 2021) 등록되지 않은 경우를 포함하면 더 많을 것이라 예상할 수 있다.

이러한 C2C 중고거래 시 일어날 수 있는 사기

거래 문제를 해결하기 위하여 다양한 방법의 연구가 진행되었다. 이들을 두 종류로 나눌 수 있는데 첫 번째는 플랫폼에서 일어나는 거래 과정 검증 및 보안 절차 등을 강화하여 사기를 해결하고자 하는 방법이다. 예를 들어 플랫폼에서 이루어지는 거래의 신뢰성 확보를 위한 방법으로 수학적 모델을 통한 신용등급 계산법이나 평판 시스템을 개선 및 추가하여 거래자 간 사기를 방지하는 방법을 개발하거나(Li and Chen, 2020; Rasheed and Olukemi, 2019), 설문을 통해 두 C2C 거래 플랫폼을 비교하여 어떤 방식의 운영이 사기 문제를 좀 더 효과적으로 해결할 수 있는지 살펴보는 연구들이 있어 왔다(Yamamoto and Ohshima, 2017). 다양한 결제 시스템 중 어떤 시스템이 사기를 막는데 더 많은 도움이 되는지 분석하거나(Zainuddin *et al.*, 2017), 블록체인을 활용하여 보안을 강화하는 방법이 모색되기도 하였다(김하정 등, 2019; 이경남, 전계형, 2018). 그러나 이 방법들의 경우 적용 과정에서 막대한 시간 및 비용이 들고 바로 적용할 수 없다는 단점이 있다. 또한 사기의 형태는 계속 변화하고 진화하기 때문에(Gupta, 2018; 이동은, 2019), 거래 과정의 보안 및 검증 절차 등을 강화하더라도 기존에 알려져 있지 않은 새로운 형태의 사기가 발생한다면 막기 어렵다는 단점이 있다. 따라서 두 번째 종류로, 사기의 근본적인 원인인 사기 행위를 하는 판매자들을 탐지하는 방법이 등장하였다(황현정 등, 2017). 그러나 휴대폰 정보와 같은 특정 카테고리 연구 대상 범위가 한정되어 있으며 게시글의 사기 여부를 판단하기보다 게시글의 신뢰도를 5단계 별점으로 도출하여 각 별점 단계가 실제 현실 세계에서 어느 정도의 신뢰도를 의미하는지 모호하다는 한계를 가진다. 또한 사람이 게시글을 직접 수집하였기 때문에 샘플 수가 적어 연구 결과의 일반화 한계 또한 가지고 있다(황현정 등, 2017). 이처럼 C2C 중고거래의 폭발적인 증가 추세에도 불구하고 거래가 실제 일어나기 전 드러나는 요소의 일반화를 통해 사기 의도 자체를 탐지하는 모델에

대한 연구는 미미한 실정이다.

C2C 중고거래는 판매자가 구매자에게 상품에 대해 게시글로 직접 설명하는 형태로 이루어지기 때문에 구매자는 게시글에 전적으로 의존할 수밖에 없으며, 구매자가 얻는 정보 원천 중 가장 큰 비중을 차지하는 것 역시 게시글이다. 게시글에는 상품에 대해 판매자가 강조하는 내용, 판매자의 게시글 작성 습관 등 다양한 정보가 포함되어 있으며 그 형태 또한 텍스트, 이미지 등으로 다양하다. 따라서 중고 거래를 가능하게 하는 기본 요소인 게시글을 대상으로 사기 탐지 모델을 개발한다면, 거래 이전 사기 요소를 탐지하는 보편화된 모델을 만들 수 있을 것이다.

이를 위해 본 연구에서는 게시글의 특성을 어떻게 파악할지에 초점을 맞추었고 국내 C2C 중고거래 플랫폼 중 가장 큰 규모인 ‘중고나라’에서 수집한 게시글 145,536건을 분석하였다. 이를 통해 일반 게시글과 사기 게시글을 비교하여 사기 게시글이 가지는 특성(Feature)을 추출하고 이를 통해 사기 게시글을 판별하는 머신 러닝(Machine Learning) 모델을 제안하고자 한다.

본 연구에서 제안하는 모델을 통해 일반 게시글과 대비되어 나타나는 사기 게시글의 특성을 파악할 수 있을 것이고, 따라서 C2C 중고 거래를 하고자 하는 구매자와 플랫폼 관리자에게 사기를 구별

하는데 도움을 주는 한편 플랫폼 내의 사기 행위 예방에 긍정적인 영향을 줄 것으로 기대한다.

## II. 문헌 연구 및 이론적 배경

### 2.1 온라인 거래 사기

온라인 거래 사기는 오래전부터 인터넷 사기의 절반을 차지할 정도로 심각한 문제였으며 <표 1>에서 나열한 것처럼 그 종류가 다양하다(Chua and Wareham, 2004). 그러나 Shilling, Bid shielding, Fee stacking, Shell auction은 인터넷 경매와 연결된 것으로 경매 방식으로 거래되지 않는 C2C 중고 거래와는 관련이 없다. 또한 C2C 중고 거래는 앞에서도 언급하였듯이 판매자가 구매자에 비해 많은 정보를 가지고 정보를 통제하며, 그로 인해 구매자의 피해가 발생하는 상황이다. 따라서 구매자의 사기에 해당하는 Failure to pay, Buy and switch, Loss or damage claims은 C2C 중고 거래 사기에서 주로 다루지는 유형은 아니다. 또한 Triangulation/fencing은 제품의 획득 원천과 관련된 내용으로 C2C중고 거래 자체의 사기라고 보기는 어렵다. 따라서 본 연구에서 다루는 사기는 판매자가 자신의 이익을 위해 상품의 설명을 사실과 다르게 적시하거나 다른 물품을 발송하는 Misrepresentation

<표 1> 인터넷 거래에서 일어나는 사기의 종류(Chua & Wareham, 2004)

구분	사기 종류	설명
경매	Shilling	판매자가 상품 가격을 높이기 위해 자신의 경매에 참여
	Bid shielding	두 명이 입찰에 참여하여 낮은 가격으로 입찰을 받기
	Fee stacking	판매자가 경매 후 부가 비용을 부과하여 높은 금액을 수취
	Shell auction	판매자가 이름과 신용카드를 얻기 위해 경매함
구매자 사기	Failure to pay	구매자가 송금하지 않음
	Buy and switch	구매자가 상품 수취 후 결함이 있는 물품을 반품
	Loss or damage claims	구매자가 상품에 결함이 있다고 주장하며 환불을 요구
제품의 획득 원천	Triangulation/fencing	판매자가 도난품을 판매
판매자 사기	Misrepresentation	판매자가 상품 설명을 의도적으로 정확하지 않게 설명
	Failure to ship	판매자가 상품을 발송하지 않음
	Reproductions and counterfeits	판매자가 위조품을 광고

사기와 판매자가 의도적으로 발송을 하지 않는 Failure to ship 사기, 그리고 위조품을 진품이라 속여 광고하는 Reproductions and counterfeits 사기라 할 수 있다.

## 2.2 사기 탐지 연구

사기 탐지 연구는 불법적인 행위를 탐지하는 방법을 찾는 연구이다. <표 2>에서 볼 수 있듯이 이미 분석회계, 허위 청구, 위조, 해킹, 탈세 등 다양한 분야에서 사용되어 사기 행위를 탐지해 왔다.

문헌들을 살펴보면 먼저 사기 탐지 연구 전반에서 대량의 실제 데이터를 사용하여 머신 러닝 방법을 이용한 연구들이 증가하고 있음을 알 수 있다. 그 예로, Fanning *et al.*(1995)의 연구에서는 382개의 기업의 재무제표를 대상으로 Generalized Adaptive Neural Network Architectures(GANNA) 모델과 Adaptive Logic Network(ALN) 모델을 적용하여 사기 재무제표를 탐지하였으며 Lin *et al.*(2003)는 기업의 재무제표 200개에 로지스틱(Logistic) 모델과 퍼지 신경망(FNN; Fuzzy neural network) 모델을 적용하여 사기 재무제표를 탐지하는 연구를 하였다. Little *et al.*(2002)의 연구는 미국이 운용하는 국가 작물 보험에서 2000년에 청구된 약 백만개의 보험을 대상으로 로그 선형(Log linear) 모델에 적용하여 다수의 허위 청구를 탐지하였으며 Wang and Xu(2018)는 자동차 보험에서 발생한 37,082건의 청구내용에 Support Vector Machine(SVM), 랜덤포레스트(RF; Random Forest), 딥 러닝(Deep Learning)을 적용하여 허위 청구를 탐지하는 연구를 하였다. Ghosh and Reilly(1994)의 연구는 발생 거래량, 거래 발생 일자 및 시간 등의 내역을 신경망에 학습시켜 두 시간마다 위험 점수를 도출하여 위조된 신용카드를 탐지하였다. 또한 Maes *et al.*(2002)은 신용카드 내역 데이터에 Bayesian Networks 모델을 적용하여 카드의 사용자가 동의한 금액보다 더 큰 금액을 사용자 몰래 청구하는 사기를 탐지하였다. Barse *et al.*(2003)은

VoD(Video on Demand) 서비스의 로그 데이터에 신경망을 활용하여 셋톱박스를 해킹하여 서비스를 사용하거나 실사용자 몰래 영화를 구매하고 해당 영화를 재배포하는 등의 범죄행위를 해결하려 하였다. de Roux *et al.*(2018)은 건축 분야에서의 탈세 범죄를 탐지하고자 1,367개의 세금신고서에서 건축 면적, 프로젝트 비용 등의 데이터를 바탕으로 클러스터링 기법을 사용하여 분포도를 작성하고 그 분포도에서 이상치라고 판단되는 데이터를 세금신고가 정상적이지 않은 건축 프로젝트로 의심할 수 있다고 하였다.

C2C 거래 분야에서도 거래를 통해 부당한 이익을 취하려는 사기 행위를 탐지하고자 하는 연구가 전체 사기 탐지 연구 중 15% 정도를 차지할 정도로 적지 않은 연구가 이루어지고 있음을 볼 수 있다 (Sinayobye *et al.*, 2018). 이들 또한 머신 러닝 방법을 주로 이용하였는데, Yoshida and Ohwada(2010)는 C2C 온라인 경매 시스템을 이용하여 부당한 이익을 취하는 사람들을 막기 위한 연구를 진행하였으며 새로운 사기 유형을 찾아내기 위해서는 머신러닝 기법 중에서 비지도 학습(Unsupervised learning)의 필요성을 제시하면서 단일 클래스 SVM 방법으로 특이한 데이터를 구별하고 이를 의사결정나무(Decision Tree)로 분석하는 방법을 제안했다. Ford *et al.*(2013)은 eBay에서 진행된 플레이스테이션3 경매 데이터를 대상으로 입찰자와 관련된 속성(attributes), 입찰의 단계에 관련된 속성, 그리고 해당 경매에 관련된 속성으로 특성을 유형 및 발견 원천에 따라 3개로 구분하였다. 그리고 이 데이터셋을 신경망에 학습시켜 이상 행동을 보이는 입찰자를 탐지하는 연구를 하였다. 또한 Tsang *et al.*(2014)은 C2C 경매 거래에서 사기 거래 데이터량이 정상 거래 데이터량에 비해 1:180의 비율로 현저히 적다는 것을 해결하기 위하여 샘플링(sampling) 기법을 사용, 1:1로 그 비율을 조정하여 머신 러닝에 사용, 탐지 정확도를 향상시키는 연구를 하였다.

위의 연구들을 통해 새로운 사기 유형을 탐지하기 위하여 비지도 학습을 포함한 머신 러닝 방법

〈표 2〉 사기 탐지 연구

탐지행위	사용 모델	연구자
분식회계	GANNA, ALN	Fanning <i>et al.</i> (1995)
	Logistic, FNN	Lin <i>et al.</i> (2003)
보험금 허위청구	Log linear,	Little <i>et al.</i> (2002)
	SVM, RF, 딥 러닝	Wang and Xu(2018)
신용카드 위조사용	신경망	Ghosh and Reilly(1994)
	Bayesian Networks	Maes <i>et al.</i> (2002)
해킹	신경망	Barse <i>et al.</i> (2003)
탈세	클러스터링	de Roux <i>et al.</i> (2018)
온라인 C2C 거래	SVM, 의사결정나무	Yoshida and Ohwada(2010)
	신경망	Ford <i>et al.</i> (2013)
	신경망	Tsang <i>et al.</i> (2014)

을 도입하는 것이 효과적일 수 있으며 거래과정 또는 행위자 별로 사기 판별에 도움이 될 것이라고 생각이 되는 특성들을 예상하고 그에 맞게 가공하는 것이 효과적일 수 있음을 알 수 있다. 마지막으로 데이터셋 자체의 구성이 사기 데이터가 정상 데이터에 비해 적을 수밖에 없는 특성을 가지므로 사기 데이터량과 정상 데이터량의 큰 비율 격차가 모델의 성능 저하로 이어질 가능성이 있고 따라서 그 비율을 조정하는 샘플링 방법이 필요하다는 것을 알 수 있다.

### 2.3 거짓말: 언어적 특성의 고려

본 연구의 목적은 C2C 중고거래에서 일어나는 사기를 억제하고 피해를 예방하는 데 기여할 수 있는 사기 탐지 모델을 제안하는 데에 있다. 따라서 C2C 중고거래의 기본 도구이자 사기 발생 전에 미리 사기글을 탐지할 수 있는 단초인 게시글을 모델 개발의 대상으로 한다. 그러기 위해서는 사기 게시글이 다른 정상 게시글과 어떤 다른 특징을 가지는지 알아볼 필요가 있다. 사기 목적으로 작성되는 게시글은 기본적으로 사실과 다른 거짓말이 들어가 있는 글이므로 본 연구의 이론적 배경으로 거짓말의 탐색부터 시작한다.

거짓말이란 ‘다른 사람에게 사전경고없이 사실이 아닌 것을 성공여부와 관련없이 의도적으로 믿게 만드는 시도’를 말한다(Vrij, 2008). 대부분의 사람들은 이러한 거짓말을 판단하기 어렵다고 알려져 있다(Bond Jr and DePaulo, 2006). 오랜 시간동안 사람들을 관리하는 직업에서 종사해오거나 범죄자 또는 거짓말을 다루는 전문가들도 거짓말을 정확히 구별하지 못하는 것으로 나타났다(Akehurst *et al.*, 1996; Ekman and O’Sullivan, 1991; Hart *et al.*, 2006). 따라서 거짓말을 보다 잘 판단하기 위한 연구들이 행해져 왔다. Zuckerman *et al.*(1981)은 이를 위해 대상에게 질문을 하고 나타나는 반응을 살펴보았는데 거짓말을 할 때는 눈 마주침, 웃음, 말하는 속도가 감소하고 자세의 변경, 대답시간, 말 더듬기, 말하는 음의 높이 등이 증가하는 것을 보였다. 또한 Vrij *et al.*(2001)은 86명의 대상자에게 영상을 보여주고 그 영상에 대해 한번은 본 것을 진술하고 한번은 본 것과 반대로 진술하는 과정에서 대상자에게 나타나는 반응을 지켜보는 실험을 진행하였다. 그 결과 거짓말을 할 때 손과 발의 움직임이 감소하였으며 대답 속도 증가, 웃음이 감소한다는 것을 알아냈다. 이와 같이 거짓말을 할 경우 인간의 신체, 행동, 언어에 어떤 현상이 나타나며, 그 현상은 시선의 움직임, 머리, 팔, 손,

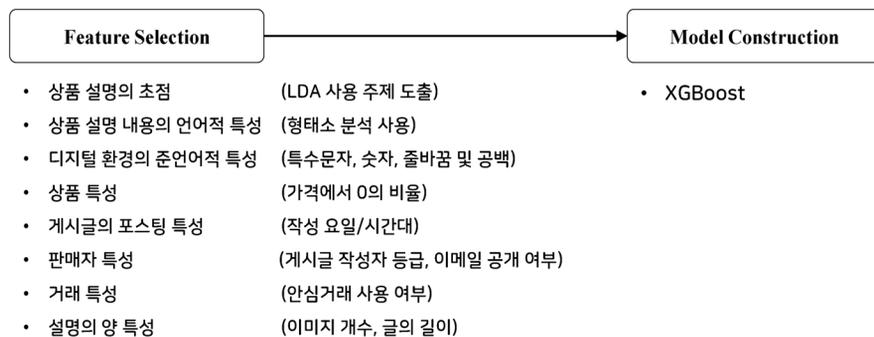
다리의 움직임, 웃음, 자세 변경, 몸짓 등의 비언어적 현상, 말 더듬기, 말하는 속도, 대답시간 등의 준언어적 현상, 대답의 길이, 짧은 문장, 논리적 모순, 상세한 설명 등의 언어적 현상으로 구분할 수 있다(이세정, 2018). 추가적으로 컴퓨터 기반 커뮤니케이션과 같은 특수한 환경에서는 느낌표, 물음표의 반복적인 사용, 말 줄임표, 공백 등이 준언어적 현상을 대체할 수 있다(Kalman and Gergle, 2010; Liebman and Gergle, 2016).

본 연구는 C2C 중고거래 플랫폼에 게시되는 글 중 사기 게시글을 판별하는 모델을 제안하고자 하므로 대면상태에서 이루어지는 비언어적 현상 또는 일부 준언어적 현상을 제외하고 컴퓨터 기반 대개 커뮤니케이션에서 발생하는 준언어적 현상 및 언어적 현상에 집중한다. 이렇게 언어적 현상을 분석하는 것은 복잡한 장비를 요하지 않는다는 장점이 있다(이석준, 2015). 또한 Smith(2001)의 연구에서는 단어의 변화, 감정표현, 대명사, 시체의 변화, 시간 등의 언어적 현상을 분석하는 SCAN (Scientific Content Analysis)라는 기법을 사용하여 70%~80% 정도의 정확도로 진술의 진위를 판별하는 결과를 얻었다. 또한 Newman *et al.*(2003)의 연구에서는 거짓 진술의 문장이 길고 복잡하다고 하였다. 반면 한국어를 대상으로 분석한 실험에서는 거짓 진술에서 문장의 길이 및 형태소의 수가 증가하였으며 명사, ‘이다’동사, 형용사 등이 증가하였고 감정 및 감각적 단어들도 증가한 것으로 나

타났다(문옥영, 2011). 이러한 연구들은 게시글의 언어적 특성을 연구 모형에 도입하는 것이 효과적일 수 있음을 시사한다.

### III. 연구모형

이론적 근거와 기존 연구를 토대로 사기 게시글 탐지를 위한 연구 모형을 구성하였다. 이를 위해 실제 사기 게시글에서 드러나는 여러 언어적, 비언어적 특성을 추출하고 이를 기반으로 머신러닝 기법을 사용하였다. 보다 구체적으로, 이 연구의 모델을 위해 사용한 특성으로, 1) LDA(Latent Dirichlet Allocation) 기법을 사용하여 상품 설명의 초점이 되는 주제를 추출하고, 2) 형태소 분석으로 단어를 추출하여 상품 설명 내용의 언어적 특성을 추출하며, 3) 디지털 환경에서 나타나는 준언어적 특성인 특수문자, 숫자, 줄 바꿈 및 공백, 4) 상품의 특성인 가격, 5) 게시글의 포스팅 적 특성인 작성 날짜, 작성 시간, 6) 판매자 특성인 게시글 작성자 등급과 이메일, 7) 거래 특성인 안심거래 사용 여부, 8) 설명의 양 특성인 글의 길이와 이미지 개수를 사용하였다. 이렇게 정리된 특성들을 머신러닝 기법 중 가장 효과적이라 알려져 있는 Gradient Boosting 기법 중 XGBoost(Chen and Guestrin, 2016)를 사용해서 사기 탐지 정확도를 높이는 최종 모델을 개발 및 제안하려 한다. 연구 모형을 도식화하여 요약하면 <그림 1>과 같다.



<그림 1> 연구 모형 구성 요약

<그림 1>의 모형을 적용하기 위하여 다음과 같은 과정을 거쳤다. 1) 데이터를 수집하여 2) 특성을 추출한 뒤 3) 그렇게 만들어진 특성을 모델이 인식할 수 있도록 전처리하고 4) 불균형 데이터의 균형을 맞춰 주기 위하여 데이터 샘플링을 하고 5) 모델에서 가장 최상의 결과를 내는 파라미터를 찾아 성능을 확인하고 의미를 해석하였다.

## IV. 데이터 수집 및 전처리

### 4.1 데이터 수집

‘Naver’가 제공하는 전자게시판 서비스인 café를 기반으로 하고 있는 ‘중고나라’는 국내에서 가장 규모가 큰 온라인 중고거래 사이트로 2021년 3월 28일 기준 1,800만 명 이상이 이용하고 있다. 해당 사이트에서는 판매자가 상품의 설명과 연락 수단을 게시해두면 구매자가 연락을 취하는 형태로 거래가 이루어지고 있다. 주로 거래가 이루어지는 게시판은 물품의 카테고리별로 분류한 게시판이고 이외에 지역별로 거래를 원하는 사람들을 위하여 따로 지역별 게시판이 존재한다. 각 게시글의 내용은 제목, 이미지, 가격, 거래방법, 배송방법 그리고 상품설명 내용 정도의 기본적인 양식이 존재하지만 상품설명의 경우 게시자가 내용의 구성 및 형태를 자유롭게 작성할 수 있다. 또한 게시글에는 게시자가 마음대로 변경할 수 없는 닉네임, 사이트 내에서의 회원등급, 이메일, 연락처와 같은 게시자의 기본정보 그리고 게시글의 속성인 게시일시가 포함되어 있다.

데이터 수집은 2020년 2월 3일부터 2020년 3월 10일까지 37일간 행해졌다. 수집일로부터 약 1년 여의 게시글을 대상으로 한 결과 2019년 2월 26일부터 2020년 2월 12일까지의 게시글이 수집되었다. Python 환경에서 Selenium을 활용하였으며 노트북, 태블릿, 데스크탑 등 카테고리별로 분류한 126개의 게시판에 올라온 게시글을 가장 최근의 게시글부터 순서대로 총 756,200건의 게시글에서

작성날짜, 작성시간, 작성자등급, 메일, 안심거래 사용여부, 가격, 상품설명의 내용, 이미지의 개수를 수집하였다.

### 4.2 데이터 레이블링

사기 게시글을 탐지하는 모델을 만들기 위해서는 먼저 사기 게시글 여부를 알 수 있는 기준이 필요하다. 이 기준은 사기 피해사례 검색 서비스를 제공하는 ‘The Cheat’ 사이트의 API를 활용하여 판단하였다. ‘The Cheat’는 온라인 거래 사기 피해자의 신고를 기반으로 운영되며 피해자가 사기글 게시자의 계좌번호와 휴대폰 번호 등을 공유하고 다른 사용자들이 이를 거래 전에 검색을 가능하게 하여 2차 피해를 예방하는 방식으로 서비스를 제공한다. 이 조회 서비스는 6개월 이내의 피해사례가 있고 계좌번호나 휴대폰번호가 있어야만 검색이 되기 때문에 6개월보다 더 이전이거나 게시글에 휴대폰번호가 기입되어 있지 않은 경우에는 사기 게시글이라도 사기 게시글 여부를 판단하기 어렵게 된다. 따라서 수집한 데이터에서 사기글 판단이 불가능한 데이터를 제외한 후 총 145,536건을 분석에 사용하였다. 이 중에서 사기 게시글로 판단된 데이터는 842건이다.

145,536건의 분석에 사용된 데이터에서 작성 시간대는 오전 11시의 글이 9,216건으로 제일 많고 오전 4시의 글이 949건으로 제일 적었다. 요일의 경우 금요일이 16,728건으로 제일 적었고 일요일이 24,703건으로 제일 많았다. 안전번호 미사용 게시글은 138,297건이며, 메일을 공개한 게시글은 138,643건이었다. 작성자는 총 31,176명이며 그 중에서 가장 낮은 등급이 26,363명으로 제일 많았고 제일 높은 등급은 6명으로 제일 적었다. 또한 이미지는 평균적으로 20.5건을 사용하였다.

### 4.3 특성 추출 및 변환

먼저 특성 추출을 위해 별도의 분석 기법이 필

〈표 3〉 변환 데이터

변환 전 데이터	변환 후 데이터
작성 날짜(예: 2020년 02월 12일)	작성 요일(예: 수요일)
작성 시간(예: 14:56)	작성 시간대(예: 14시)
작성자 등급(예: 성실등급)	작성자 등급(예: 성실등급)
메일(예: ****74@naver.com)	메일 공개 여부(예: 공개일 경우 Y, 그렇지 않을 경우 N)
가격(예: 10,000원)	가격 중 0의 비율(예: 20%)
이미지의 개수(예: 12)	이미지의 개수(예: 12)

요하지 않은 특성들을 선정하고 전처리하였다. 작성날짜, 작성시간, 작성자등급, 이메일, 안심거래 사용여부, 가격, 상품설명 내용, 이미지의 개수를 사기글을 판단하는데 사용할 수 있는 특성으로 보아 추출하였다. 이후 이들을 분석에 필요한 형태로 변환하여 사용하였으며 그 결과는 <표 3>과 같다.

작성된 일자 및 작성 시간은 사기글이 작성된 일자과 시간을 수집한 데이터이다. 사기글이 어느 요일, 그리고 시간대에 자주 작성되는지를 알아보기 위해 일자는 요일로 시간은 시간대로 변환하여 사용하였다.

작성자 등급은 방문 수와 중고나라 앱을 동시에 사용하는 등의 특정 조건을 만족할수록 높은 등급이 부여되는 ‘중고나라’ 내에서의 등급을 말한다. 사기글 게시자의 특성상 아이디를 변경하며 게시를 할 것으로 추측되어 높은 등급에 있지 못할 것으로 판단하였으며 형태 변환 없이 사용하였다.

작성자의 이메일은 게시글에 공개한 메일주소를 말한다. 메일 자체로는 사기글을 판단하는데 도움이 되지 않는다고 판단하였으나 메일을 공개한 사람과 공개하지 않은 사람 간 차이가 있을 것으로 판단되어 메일 공개 여부로 변환하였다.

가격은 작성자가 중고나라 양식의 가격란에 제시한 상품 가격을 말한다. 판매자가 자유롭게 입력하는 방식이기 때문에 실제 희망 판매가를 기재한 경우도 있지만 천원, 만원 단위를 절사한 금액을 기재하기도 하고 희망 판매가로 보이지 않는 숫자(예: 121212)를 기재하는 경우도 존재한다. 따

라서 의미 없는 숫자, 천원, 만원 단위를 절사한 금액 등을 기재하는 행위를 가격 자체에서 찾을 수 있는 특징으로 보고 이를 구별하기 위한 지표로 가격에 제시된 숫자의 총 개수에서 0의 비율을 가격 특성을 파악할 수 있는 특성으로 사용하였다. 사기 행위는 반복적으로 이루어지기 때문에 반복되는 게시글 작성에 대한 피로감, 실제 물건을 소유하고 있지 않기 때문에 세밀한 가격 제시의 어려움 등에서 정상 게시글과는 다른 차이를 보일 것으로 판단된다.

이미지의 개수는 게시자가 내용 설명란에 사용한 이미지의 개수를 의미한다. 사기 게시글과 정상 게시글 사이에 게시 이미지 수의 차이가 있을 것이라고 가정하였다.

또한 데이터 관찰 결과, 구매자의 관심을 끌고자 하는 목적의 게시글이기 때문에 특수문자, 숫자, 공백 등이 더 많이 등장한다는 점을 발견하였고 이는 컴퓨터 기반 커뮤니케이션에서 발생하는 준 언어적 현상으로 볼 수 있다(이나은, 이상원, 2017). 따라서 본 연구에서는 특수문자, 숫자, 줄바꿈 및 공백의 비율 또한 측정하였다

## V. 토픽모델링: 언어적 특성 추출

수집한 데이터 중 상품설명 내용은 언어적 특성이 반영되는 내용이므로 이를 거짓말과 관련된 문헌연구를 바탕으로 특성으로 가공하기 위해서 토픽모델링 분석 기법을 사용하였다.

먼저 분석 대상이 텍스트 형태의 데이터이므로

이러한 텍스트 형태의 데이터의 경우 언어적 현상을 측정하기 위하여 형태소 분석을 활용하였다. 한국어 형태소 분석기에는 kharii, 한나눔, 꼬꼬마, KOMORAN, OKT, mecab이 있으나 Windows 환경에서 사용이 가능하며 분석소요시간 대비 성능을 고려하여 okt 형태소 분석기를 선택하였다. 그리고 난 후 한국어에서 거짓 글을 작성할 경우 나타나는 특징으로 명사, ‘이다’동사, 형용사가 증가하고 전체 문장의 길이 등이 증가한다는 연구를 참고하여(문옥영, 2011) 전체 문장 형태소 대비 명사, 동사, 형용사, 부사의 개수와 문장 전체의 길이를 측정하였다.

다음으로, 사기를 목적으로 하는 글에서 쓰이는 단어 중 다른 글에서는 잘 쓰이지 않는 단어를 더욱 강조하기 위하여 TF-IDF(Term Frequency - Inverse Document Frequency) 값을 도출하였다. TF-IDF값은 TF(Term Frequency)와 IDF(Inverse Document Frequency)를 곱해서 산출되는 값이다. 즉, 단어의 단순한 출현 횟수에 단어가 다른 문서에서 등장하는 빈도의 역의 값을 곱하는 것이다(Jones, 1972; Luhn, 1957). 따라서 TF-IDF값은 단어의 빈도를 측정하되 다른 문서에서 잘 등장하지 않는 단어를 강조할 수 있다. 따라서 TF-IDF를 통해 특정 게시글에서 등장하는 특이한 단어를 파악하였다.

이를 기반으로, 토픽 모델링 기법인 LDA모델을 이용하여 게시글의 주제를 파악하고 이를 모델의 특성으로 사용하였다. LDA 기법에서는 연구자가 도출될 적절한 주제 수를 정해주어야 하며 이를 위해 기존 연구들이 제시한 지표로 혼잡도(Perplexity)<sup>2)</sup>와 일관성(Coherence)<sup>3)</sup>이 있다. 고전적으로 많이 쓰였던 혼잡도는 cross-entropy값을 기본으로 하는 정량적인 언어 모델 평가 방법이다

(Brown *et al.*, 1992). 다음에 출현할 것으로 예측되는 단어가 많을수록 각 단어의 확률은 줄어들고 혼잡도는 커지며 좋지 못한 언어 모델이라고 볼 수 있다. 즉 혼잡도가 작을수록 좋은 모델이다. 하지만 혼잡도는 사람이 해석하기 용이해야 한다는 측면은 고려하지 않는다는 단점이 있으며 실제로 혼잡도가 좋은 점수가 나오더라도 해석하기 더 나은 모델임을 의미하지 않는다는 연구도 존재한다(Chang *et al.*, 2009). 일관성은 이런 문제를 해결하고 사람이 해석하기에 적합한지를 알아볼 수 있는데 우선적으로 혼잡도보다 일관성을 사용하여 토픽 수를 결정하는 것이 권장된다. 일관성은 각 토픽의 상위 단어들 간의 유사도를 계산하여 실제로 의미적으로 일치하는지 판단하는 지표이다(Newman *et al.*, 2010). 단어가 같이 등장하는 확률이 높을수록 일관성은 높아지기 때문에 일관성이 높을수록 좋은 언어 모델이라고 할 수 있다.

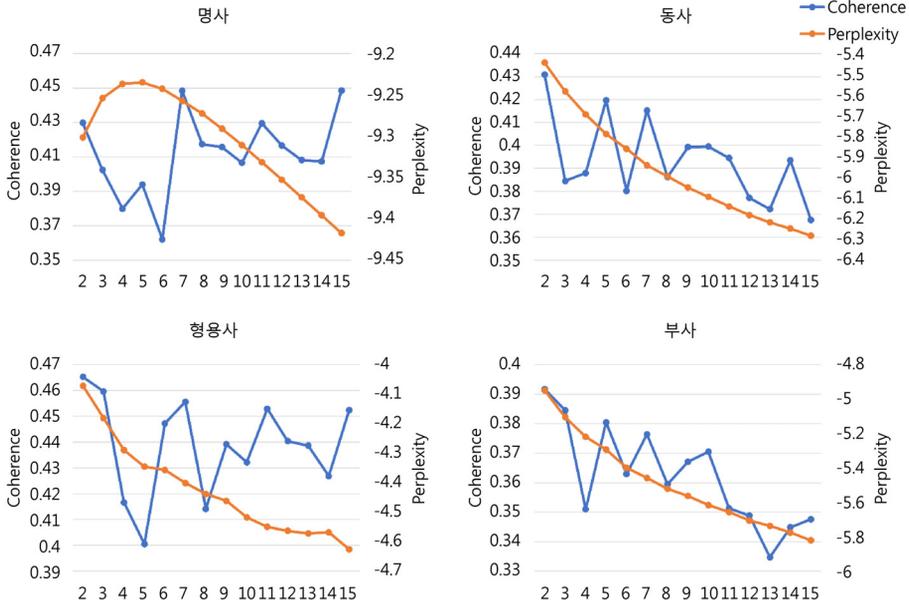
본 연구에서는 보다 상세한 주제를 파악하기 위하여 명사, 동사, 형용사, 부사의 네 가지 품사를 대상으로 각각 주제를 도출하였다. 도출된 LDA 모델에 대한 혼잡도와 일관성을 구하여 먼저 일관성을 이용하여 주제 수를 판단하고 해석이 어려울 경우 혼잡도와 함께 고려하여 주제 수를 정하였다. 그 결과 <그림 2>와 같이 도출되었으며 네 개의 품사 모두 주제수가 2일 때 가장 좋은 결과가 도출된다고 판단하였다.

<그림 2>에서 보이듯 명사를 대상으로 주제 수에 따른 일관성과 혼잡도를 파악하였더니 일관성의 경우 주제의 수가 2일 때 0.429, 7일 때 0.448로 좋은 결과를 보였으며 추가로 혼잡도까지 고려하면 주제가 2개일 때 .9.301로 작게 나와 가장 적절한 것으로 판단되었다. 두 개의 주제를 살펴보았더니 명사의 경우 상품, 사이즈, 상태, 사용감, 가격 등의 단어들이 주로 등장하여 주제1은 상품 정보들로 판단하였으며(이하 ‘명사 주제1’), 주제 2의 경우에는 판매, 연락, 직거래, 가격, 택배 등 상품 정보가 아닌 전달방법 또는 연락방법 등의 단어가 주로 등장하여 상품전달/연락 방법(이하 ‘명

2)  $Perplexity = 2^{H(W)} = 2^{-\frac{1}{n} \log_2 p(w_1, \dots, w_n)}$

(H(W)): 단어w의 집합인 문장W에 대한 cross entropy).

3)  $Coherence = \log \frac{p(w_i, w_j)}{p(w_i)p(w_j)}, i \neq j.$



〈그림 2〉 품사별 LDA모형의 일관성(Coherence)과 혼잡도(Perplexity)

사 주제 2’)으로 설정하였다.

동사의 경우 일관성을 고려하였을 때 2개의 주제일 때 0.430으로 가장 높은 값이 나와 2개의 주제가 가장 적절하다고 판단하였다. 동사의 LDA 분석 결과에서는 먼저 전체 문서에서 자주 사용되어 두 주제 모두에서 동일하게 등장한 ‘하다’를 제외하였다. 이후 ‘주다’, ‘드리다’, ‘보내다’, ‘반다’의 동사들이 주로 등장한 주제 1(이하, ‘동사 주제 1’)은 전달과 관련 있다고 판단하였으며 주제 2(이하, ‘동사 주제 2’)는 ‘입다’, ‘적다’, ‘줄다’, ‘되다’의 동사들이 등장하여 상품의 상태와 관련 있다고 판단하였다.

형용사를 대상으로 한 LDA 분석에서 일관성이 가장 높은 경우는 주제의 수가 2인 경우 0.465로, 따라서 주제의 수를 2로 설정하였다. LDA 분석 결과 두 주제 모두에서 동일하게 등장한 ‘이다’를 제외하고 ‘좋은’, ‘예쁜’, ‘이쁜’, ‘고급스러운’의 형용사가 주로 등장하는 주제 1은 상품에 대한 평가(이하, ‘형용사 주제 1’)라고 판단하였으며 ‘가능한’, ‘부탁드리는’, ‘원하는’, ‘신중한’이 등장한

주제 2는 행위와 관련 주제(이하, ‘형용사 주제 2’)로 구분하였다.

부사를 대상으로 한 LDA 분석에서의 주제 수 또한 주제가 2일 경우 일관성이 0.391로 가장 높게 나타나 주제를 2개일 설정하는 경우가 가장 적합한 것으로 판단되었다. 부사의 경우 두 주제 모두에 나타난 ‘다’, ‘함께’를 제외하고 ‘너무’, ‘엄청’, ‘훨씬’과 같이 어떤 형용사를 강조하는 부사가 주제1(이하, ‘부사 주제 1’)에서 주로 나타났으며 ‘없이’, ‘많이’, ‘따로’의 상태를 나타내는 부사가 주제2(이하, ‘부사 주제 2’)로 구분되었다.

이렇게 각 품사별로 2개의 주제를 추출하였으며 게시글마다 해당 주제에 해당할 확률을 도출할 수 있다. 1에 가까울수록 주제 1에 가까우며 0에 가까울수록 주제 2에 가까운 것을 의미하며 이 수치를 특성으로 사용하였다.

## VI. 모델 구축 및 평가

모델 구축 및 평가를 위하여 XGBoost를 사용하

〈표 4〉 모델에 사용되는 특성 변수

특성	설명
작성 요일	게시글이 작성된 요일(일요일 ~ 토요일)
작성 시간대	게시글이 작성된 시간대(24개의 시간대)
작성자 등급	‘중고나라’에서 작성자의 등급(5개의 등급과 등급 없음)
메일 공개여부	메일 주소를 게시글에 공개하였는지 여부(yes/no)
안전번호 사용여부	안전번호를 사용하였는지 여부(yes/no)
이미지의 개수	상품 설명에 포함된 이미지의 개수
가격 중 0의 비율	가격에서 ‘0’이 차지하는 비율
텍스트의 길이	상품 설명의 텍스트 길이
특수 문자 비율	상품 설명의 텍스트 길이 대비 특수 문자의 비율
공백 비율	상품 설명의 텍스트 길이 대비 공백의 비율
숫자 비율	상품 설명의 텍스트 길이 대비 숫자의 비율
명사 비율	상품 설명의 형태소의 수 대비 명사의 비율
동사 비율	상품 설명의 형태소의 수 대비 동사의 비율
형용사 비율	상품 설명의 형태소의 수 대비 형용사의 비율
부사 비율	상품 설명의 형태소의 수 대비 부사의 비율
명사 주제	사용된 명사들의 주제(주제1: 상품 정보, 주제2: 상품 전달/연락 방법)
동사 주제	사용된 동사들의 주제(주제1: 전달, 주제2: 상품 상태)
형용사 주제	사용된 형용사들의 주제(주제 1: 상품 평가, 주제 2: 행위 관련)
부사 주제	사용된 부사들의 주제(주제 1: 강조, 주제2: 상태)

였다. XGBoost는 의사결정나무를 부스팅 방식으로 앙상블(ensemble)하는 그래디언트 부스팅(Gradient Boosting) 방법을 기반으로 한 모델이며 병렬 계산을 통해 일반 그래디언트 부스팅 모델에 비하여 10배 이상의 계산 속도를 보이는 모델이다. 또한 머신러닝 알고리즘에서 가장 좋은 성능을 보이고 있다고 알려져 있다(Chen and Guestrin, 2016).

XGBoost에 입력하기 위해 토픽모델링과 데이터 전처리 과정에서 선정된 특성은 <표 4>와 같다. 작성 시간대와 이미지의 개수 특성은 황현정 등(2017)의 연구에서 기인하였으며 텍스트의 길이, 특수 문자, 공백, 숫자의 비율 특성은 Vrij *et al.*(2001)과 이세정(2018)의 준 언어적 특성이 거짓 탐지에 유용하다는 연구와 Kalman *et al.*(2010)과 Liebman *et al.*(2016)의 컴퓨터 기반 커뮤니케이션에서는 준언어적 현상이 다른 방식으로 나타난다는 연구들을 토대로 고안하였다. 명사, 동사, 형용

사, 부사의 비율 특성은 거짓 글이 정상 글과 사용하는 품사의 비율이 다르다는 문옥영(2011)과 Smith(2001)의 연구를 참고하였다. 또한 명사, 동사, 형용사, 부사로 주제 특성을 사용한 것은 자동차 보험 사기 탐지 연구에 주제를 특성으로 사용한 Wang and Xu(2018)의 연구를 참고하여 고안하였다.

범주형 특성인 ‘작성 요일’, ‘작성 시간대’, ‘작성자 등급’, ‘메일 공개 여부’, ‘안전번호 사용여부’는 원 핫 인코딩 방식을 적용하여 숫자로 변환하였다. 수치형 변수의 경우는 각 특성 간의 영향력의 차이가 발생할 수 있기 때문에 Scaler를 적용하는데 이상치가 존재할 수 있다는 점을 고려하여 Robust Scaler를 사용하여 표준화하였다.

## 6.1 데이터 샘플링

사기를 탐지하는 모델의 경우 사기 관련 데이터

가 정상적인 데이터보다 현저히 적다는 특징이 있기 때문에 불균형 데이터셋에 대한 처리가 반드시 필요하다(Han et al., 2005). 본 연구의 경우에도 전체 데이터셋에서 25%의 테스트 데이터를 제외한 109,152건의 훈련 데이터에서 622건만이 사기 게시글로 클래스 간 데이터 비율의 격차가 심한 불균형 데이터이다. 이를 해결하는 동시에 데이터의 치명적 손실을 피하고 범주형 변수가 있어도 사용 가능한 SMOTE-NC 오버샘플링 방식을 사용하였으며 10:3의 비율로 조정하였다(Han et al., 2005).

### 6.2 특성 중요도 파악

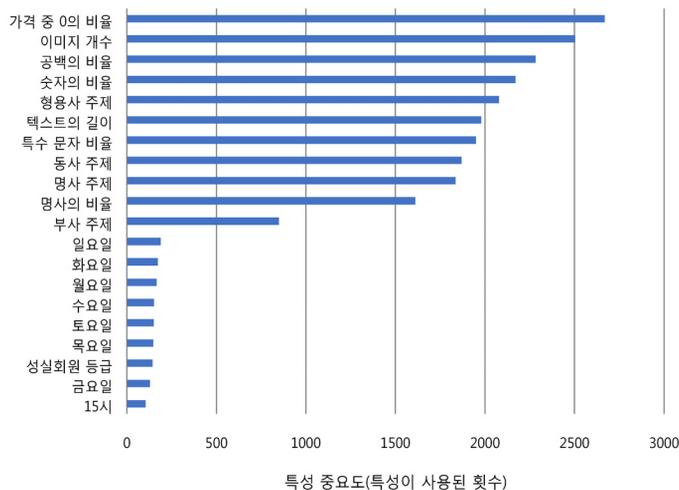
모델 개발을 위해 먼저 매개변수를 조정하였다. 학습률, 최대 나무 깊이, 부스팅 횟수 파라미터를 조정하여 최적의 성능을 보이는 파라미터 값을 선택하였으며 그 결과 학습률은 0.3, 최대 나무 깊이는 20, 부스팅 횟수는 5,000번 부스팅 하도록 하였으며 50번 이상 성능의 향상이 없다면 중지하도록 하였다. 이 과정을 거친 후 모델에서 사용한 여러 특성 중 어떠한 특성이 사기글을 판별하는데 중요한지 파악하기 위하여 XGBoost 분석 결과 확인되는 특성 중요도를 살펴보았다.

<그림 3>은 XGBoost 모델에서 사기 게시글을 판단하는데 중요하게 작용한 특성 중 상위 20개 특성의 중요도를 보여준다. 특성의 중요도가 높을 수록 사기 게시글을 판별할 때 유용한 정보가 된다. 사기 게시글을 판별할 때 글을 작성한 시간대, 요일, 등급과 같이 포스팅의 특성, 거래의 특성이나, 판매자의 특성을 보는 것보다 작성된 상품에 대한 설명 부분에서 명사의 비율과 같은 언어적 특성, 공백, 숫자, 특수문자의 비율과 같은 준언어적 특성, 주제 등을 살펴보는 것이 더 유용하다는 것을 알 수 있다.

또한 사기 게시글이 게시되는 시간대가 사기 게시글을 판단하는데 도움이 된다는 연구결과(황현정 등, 2017)와 달리 시간대는 대부분 특성 중요도에 있어 20위권 밖이며 요일 또한 사기 게시글을 판단하는데 상대적으로 역할이 미미하다는 것을 알 수 있다.

### 6.3 최종 특성 선택

앞서 살펴보았듯이 모델에 사용된 특성 모두가 사기글 탐지를 보다 잘 하는데 긍정적인 영향을 주는 것은 아니며 기여하는 정도도 모두 다름을



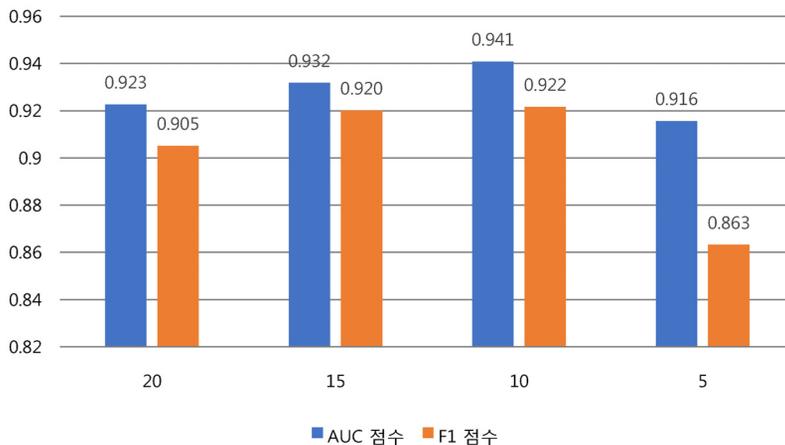
<그림 3> XGBoost 모델의 특성 중요도

확인할 수 있었다. 따라서 본 연구에서는 모델 성능을 높여주는 특성을 선택하는 과정을 거쳐 사기 글을 탐지하는 최종 모델을 확정하고자 특성의 수를 5, 10, 15, 20개로 변경해가며 그 결과를 비교하였다. 이 때 최대 나무 깊이는 특성을 몇 번 분류할지 결정하는 것으로써 특성의 수에 따라 결정해야 하는데 너무 깊다면 과대적합의 문제가, 너무 얕다면 과소적합의 문제가 발생하기 때문에 최대 나무 깊이 또한 조정이 필요하다. 동등한 조건을 위해 최대 나무 깊이를 조정하여 가장 성능이 좋은 경우를 탐색하였다. 그 결과 특성의 수가 20개일 경우에는 12, 15개일 경우에는 11, 10개일 경우에는 7, 5개일 경우에는 5로 각 모델에 대한 최대 나무 깊이를 결정하였다. 각 모델의 평가는 AUC 점수, F1 점수를 통해 비교하였다. AUC 점수는 실제로 True인 데이터 중 모델이 True로 판단한 비율(True Positive Rate)과 실제로 False인 데이터 중 모델이 True로 판단한 비율(False Positive Rate) 간의 모든 임계값에서 성능을 보여주는 ROC 그래프의 면적을 의미하며 1에 가까울수록 모델의 성능이 좋다는 것을 의미한다. 또한 F1 점수는 모델이 True인 데이터 중 실제로 True인 비율(Precision)과 실제로 True인 데이터 중 모델이 True로 판단한 비율(Recall)의 조화평균을 구한 값이다. <그림 4>

에서 보듯이 AUC 점수와 F1 점수가 0.940, 0.922로 특성의 수가 15개일 때 모델의 성능이 가장 좋았고 10개일 때 AUC 점수와 F1 점수가 0.931, 0.920으로 다음으로 높았으며 그 뒤로 20개일 때, 5개일 때 순으로 높은 성능을 보였다. 따라서 AUC 점수와 F1 점수를 비교해 보았을 경우 특성의 수가 10개일 때 최상의 결과를 보여주어 특성을 10개로 하는 모델을 최종 모델로 확정하였다. 따라서 가격 중 0의 비율, 이미지 개수, 공백의 비율, 숫자의 비율, 형용사 주제, 텍스트의 길이, 특수문자의 비율, 동사 주제, 명사 주제, 명사의 비율까지 총 10개의 특성이 최종 선택되었다.

#### 6.4 최종 분석 결과

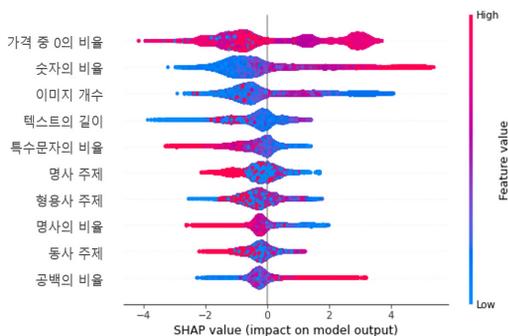
의사결정나무를 기반으로 하는 앙상블 모델의 경우 특성마다의 중요도를 볼 수 있으나 그 특성이 어떤 방향으로 영향을 미치는지는 알 수 없다(Štrumbelj and Kononenko, 2014). 하지만 SHAP(Shapley Additive exPlanation) 방법을 활용한다면 해석이 가능할 수 있다(Štrumbelj and Kononenko, 2014). SHAP은 협조적 게임이론의 Shapley value를 기반으로 한다(Shapley, 1951). Shapley value는 특정 데이터 포인트에서 해당 특성이 모델의 예측값에 기여한



<그림 4> 특성의 수에 따른 AUC, F1 점수

정도를 구하는 것으로 모든 경우의 수에서 해당 특성이 존재할 경우의 예측 값과 존재하지 않을 경우의 예측값의 차이인 한계 기여분을 구하여 가중 평균한 값을 말한다. 즉, Shapley value는 해당 데이터포인트에서 가지는 특성의 영향력이라고 볼 수 있다. 의사결정나무를 기반으로 하는 앙상블 모델에 Shapley value를 적용할 수 있게 만든 것이 TreeSHAP이며 이를 SHAP 라이브러리가 제공한다 (Lundberg and Lee, 2017).

최종 모델 결정 후 <그림 5>와 같이 SHAP Summary Plot을 통하여 사기 게시글을 탐지하는 데 특성의 값에 따른 특성의 영향력을 볼 수 있다. 각각의 데이터를 모두 점으로 표현한 것이기 때문에 두꺼운 부분은 그만큼 데이터가 그 곳에 밀집되어 있다는 의미이며 이 때 색이 붉은 점은 특성 값이 높은 데이터 포인트이며, 파란색은 특성 값이 낮은 데이터 포인트를 뜻한다. 또한 Shapley value값이 높으면 높은 예측값을, 낮으면 낮은 예측값을 예측하는데 기여했다는 것을 의미한다. 즉 본 모델에서는 사기글을 1, 정상글을 0으로 처리했으므로 Shapley value가 높은 지점에 위치한 특성의 데이터일수록 사기글을 예측하는데 기여하고 반대로 Shapley value가 낮은 지점에 위치한 특성의 데이터일수록 정상글을 예측하는데 기여했다고 볼 수 있다.



<그림 5> 특성 별 데이터 포인트의 SHAP 값

<그림 5>를 살펴보면 특성들을 두 종류로 나눌

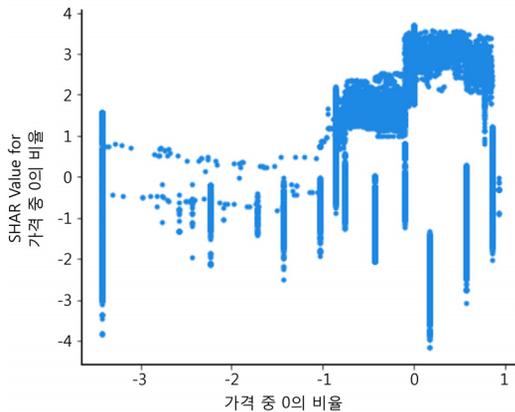
수 있다. 첫째, 숫자의 비율, 특수문자의 비율, 명사의 비율, 공백의 비율, 게시글의 주제 중 명사 주제, 동사 주제, 형용사 주제와 같이 SHAP Summary Plot 상에서 높은 특성 값과 낮은 특성 값의 위치가 대체로 분명히 나뉘는 특성들이다. 반대로 두번째 종류는 가격 중 0의 비율, 이미지 개수, 텍스트의 길이와 같이 SHAP Summary Plot 상에서 높은 특성 값과 낮은 특성 값이 혼재되어 분포되어 있는 특성들이다. 후자의 경우 해석을 위해 의존도(Dependence plot) 분석이 추가로 필요하므로 먼저 전자의 경우에 속하는 특성들에 대한 결과부터 해석하면 다음과 같다.

숫자의 비율의 경우 이것이 높은 경우는 사기 게시글을, 반대의 경우는 정상 게시글인 경향이 있다는 것을 알 수 있다. 이는 전체 글의 길이에서 숫자가 차지하는 비율이기 때문에 숫자의 비율이 높다는 것은 단순히 숫자를 많이 쓴다는 것이 아니라 글에 비해 숫자를 많이 쓴다는 의미이며 주목할 부분은 Shapley value가 높은 경우에는 넓게 퍼져 있는 반면에 Shapley value가 낮은 경우에는 숫자의 비율이 낮은 쪽으로 밀집해 있다는 점이다. 즉, 정상 게시글은 거의 대부분 숫자보다 글을 많이 쓴다는 점을 알 수 있다. 특수문자 비율의 경우는 Shapley value가 양수일 때 낮은 특성 값을 Shapley value가 주로 높은 특성 값이 넓게 분포된 것으로 보아 사기 게시글의 경우는 특수문자 비율이 상대적으로 작은 경향을 보인다고 할 수 있다. 명사의 비율의 경우는 Shapley value가 높을 때 명사의 비율이 낮으며 Shapley value가 낮을 때 명사의 비율이 높은 것을 알 수 있다. 즉, 사기 게시글 일수록 명사의 사용이 적다는 것을 알 수 있다. 공백의 비율은 Shapley value가 높을수록 공백의 비율이 높고 Shapley value가 낮을수록 공백의 비율 또한 낮다. 즉, 사기 게시글 일수록 공백, 줄바꿈 등이 잦다는 것을 알 수 있다.

주제에 대한 특성은 두 개의 주제 중 하나의 값을 갖게 되므로 그 특성 값이 0과 1사이의 값이며 1에 가까울수록 해당 글이 해당 주제와 가깝다

는 의미이다. <그림 5>에서 보이듯 명사, 동사, 형용사 주제 모두 Shapley value가 낮을 때, 즉 사기 게시글로 예측되는 경향이 있을 때는 붉은색, 푸른색 데이터포인트가 모두 존재하므로 특정 주제와의 연관성을 찾기 어려우나 Shapley value가 높을 때, 즉 정상 게시글로 예측되는 경향이 있을 때에는 명사에서는 붉은색, 동사에서는 붉은색, 형용사에서는 푸른색 데이터포인트로 뚜렷하게 나타나 특정주제와 연관성이 높다는 것을 보여준다. 정리하면 정상 게시글일수록 명사에서는 상품 정보, 동사에서는 전달, 형용사에서는 행위와 관련된 단어들이 등장하였다.

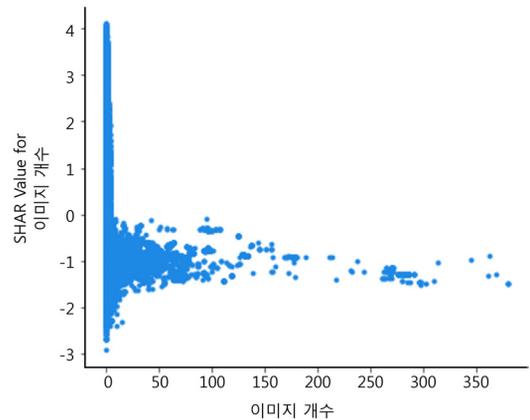
의존도를 통해 추가로 살펴본 특성들의 해석은 다음과 같다. 가격 중 0의 비율의 경우 <그림 6>을 보면 Shapley value가 음수인 부분에서의 가격 중 0의 비율은 전반적으로 낮게 분포하는데 비해 Shapley value가 양수인 부분, 즉 사기 게시글을 판단하는데 기여가 큰 부분에서는 높은 0의 비율이 존재하는 것을 볼 수 있다. 즉, 가격 중 0의 비율이 높다면 사기 게시글로 의심해 볼 수 있다는 것이다.



<그림 6> '가격 중 0의 비율' 특성의 의존도

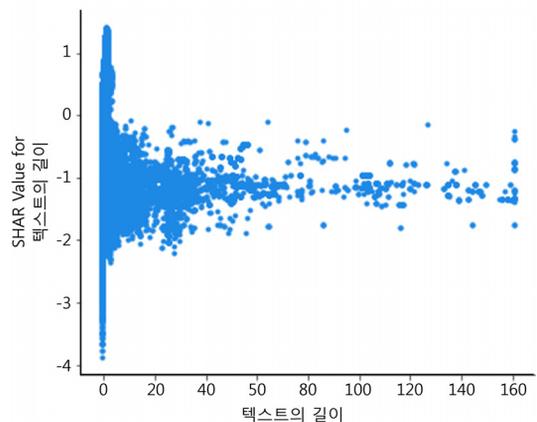
이미지 개수의 경우 <그림 7>에서 볼 수 있듯, Shapley value가 양수일 때 이미지 개수가 적다는 것을 알 수 있다. 반면 Shapley value가 음수일 경우에는 이미지 개수가 적은 것도 있지만 많은 것도

있는 등 이미지 개수의 분포가 넓게 퍼져 있다. 즉, 정상 게시글에는 이미지의 개수가 많거나 적을 수 있지만, 사기 게시글에 사용되는 이미지의 개수는 대체로 적다는 것을 알 수 있다.



<그림 7> '이미지 개수' 특성의 의존도

텍스트의 길이는 <그림 8>에서 보이듯 Shapley value가 양수일 때 텍스트의 길이가 짧은 것으로 나타났으며 Shapley value가 음수일 때는 넓게 분포한다는 것을 알 수 있다. 이 결과는 이미지 개수에 대한 분석과 유사하게 정상 게시글은 텍스트의 길이가 매우 다양하지만 사기 게시글의 경우에는 텍스트의 길이가 상당히 짧다는 것을 알 수 있다.



<그림 8> '텍스트의 길이' 특성의 의존도

## VII. 결 론

### 7.1 분석 결과 논의

분석 결과 사기 게시글은 글이 짧고, 이미지의 수가 적은 등 사기 행위자들은 게시글을 작성할 때 정상 게시글보다 대체로 상품에 대한 정보를 적게 제공하는 특성이 있는 것으로 나타났다. 이는 이들이 정상적인 상품판매를 목적으로 하지 않으므로 정확하고 상세한 정보 자체를 제공하는 것이 어렵거나 불가능하기 때문으로 보인다. 또한 대체로 사기 행위를 반복하여 시도하기 때문에 정보를 적게 제공한 처음의 시도가 성공했다면 이런 경향은 사기 행위가 반복되면서 더욱 심해질 것으로 추측해볼 수 있다. 또한 글의 길이에 비해 숫자, 공백, 줄 바꿈을 많이 사용한다는 것을 알 수 있는데 이는 글의 길이의 해석과 일맥상통한다고 볼 수 있다. 즉, 제공할 정보 자체가 적으므로 정말 필요한 정보인 가격이나 연락처 등의 숫자나 공백으로 포스팅의 길이를 늘리는 것으로 보인다. 이렇게 구체적인 내용 없이 글의 길이만 늘린다는 결과로부터 재미있는 추측을 할 수 있는데, 온라인 포스팅으로 이루어지는 사기의 경우 사기글 게시자들이 실제 내용을 채워 넣는 노력을 화면 스크롤을 늘리는 노력으로 어느 정도 대체할 수 있을 거라 생각한다는 것이다. 오프라인에서 거짓을 기술하는 종이의 여백과 빈 종이의 수를 일부러 늘리지는 않는다는 점을 고려할 때 이러한 점은 사기의 수단이 온라인으로 구현되며 발생하는 특이한 점이라 할 수 있다.

명사의 비중이 낮은 것과 연결 지어 해석한다면 명사가 적은 짧은 글이 많다는 해석도 가능하다. 구체적인 상품 정보를 제공하려면 상품과 관련한 특징들을 기술해야 하는데 명사가 적다는 특징은 역시 이러한 기술을 제대로 하지 못하고 있다는 증거가 될 수 있을 것이다. 이는 진실한 진술이 거짓 진술에 비해 복잡한 문장을 사용한다는 Newman *et al.*(2003)의 연구결과와 일맥상통하고,

거짓진술에서 단어와 문장의 길이가 증가하였다는 문옥영(2011)의 연구 결과와는 정반대의 결과를 보인다. 과거의 거짓 행위에 대한 기술이 아니라 미래에 발생시키려는 거짓에 대한 기술이라는 점과 온라인이라는 환경을 고려했을 때는 후자의 연구보다 전자의 연구가 더 잘 맞는다는 것을 보여주는 결과이다. 또한 특수문자의 사용은 일반적으로 구매자들의 눈에 잘 띄게 하기 위해서 사용하는 경우가 많았는데 사기 행위자는 잘 사용하지 않는 것으로 보아 판매가 목적이 아니기 때문에 구매자들에게 잘 보이려는 노력을 하지 않는 것으로 보인다. 이 또한 이미 저지른 과거의 거짓을 방어하는 것이 아니므로 보다 비적극적인 태도를 취하는 것으로 생각할 수 있다. 이미지의 개수에 대해서는 사기 행위자가 글을 작성할 때 이미지를 거의 쓰지 않는다는 것을 알 수 있다. 상품을 실제로 가지고 있지 않거나 기술한 내용과 같지 않은 상태의 상품을 가지고 있는 것에서 기인한다고 추측할 수 있다.

또한 사기 행위자가 게시한 글에서 가격에는 0의 비중이 많은 편이다. 예를 들어 정상 게시글에 비해 천 단위 등 상세한 수준으로 가격을 작성하기보다는 십만 단위 등으로 가격을 작성한다고 볼 수 있다. 일견 사기글 게시자의 경우 보다 더 매력 있어 보이는 가격으로 조정된 후 구매 구매자를 유인하는 노력을 할 것으로 추측할 수 있는데 본 연구 결과는 그 반대의 결과를 보여주었다. 추측컨데 실제 물건을 소유하고 있지 않기 때문에 세밀한 가격 제시에 어려움이 있는 것이라고 보인다. 또한 판매를 원하는 최종 가격을 고심 끝에 적기보다는 관심 있는 구매자가 나타나면 가격 조정을 얼마든지 해서라도 거래를 유도하면 되므로 정상 게시글 사용자보다 가격을 상세한 수준으로 적지 않는 것이라고도 생각해 볼 수 있다. 또한 사기를 대상으로 하는 품목의 가격이 대체로 비싸다는 추측도 가능하다. 나아가서는 상시 게시되어 있는 온라인 게시글의 특성과 얼마든지 다른 사기 게시글을 게시할 수 있다는 특성 상 온라인 중고

거래 환경에서의 사기글 게시자에게는 구매자를 유인하여야 하는 점이 상대적으로 어려운 요소가 아닐 수도 있다고 추측해볼 수 있다.

추가로, 사기 게시글이 게시되는 시간대가 사기 게시글을 판단하는데 도움이 된다는 연구결과(황현정 등, 2017)와 달리 시간대는 대부분 특성 중요도에 있어 20위권 밖이며 요일 또한 사기 게시글을 판단하는데 상대적으로 역할이 미미하다는 것을 보였다.

주제와 관련하여 정상 게시글은 명사의 경우 상품정보, 동사의 경우 전달, 형용사의 경우 행위와 관련된 단어들을 많이 사용한 것에 비해 사기 게시글은 뚜렷한 주제를 가지지 않는 것으로 나타났다. 이는 거짓 진술을 대답의 길이, 짧은 문장, 논리적 모순, 상세한 설명 등의 언어적 현상으로 구분할 수 있다(이세정, 2018)는 연구와 관련하여 논리적 모순을 방지하기 위하여 상세한 설명을 포기하는 것으로 생각해볼 수 있다.

## 7.2 한계점 및 향후 연구

본 연구는 다음과 같은 한계점을 가진다. 첫째, 본 연구에서는 XGBoost 모델을 사용하였으나 딥러닝과 같이 정확도 측면에서 더 좋은 모델이 있을 수 있다. 그러나 본 연구가 목표하는 바가 사기 게시글을 탐지하는데 도움이 되는 실제적 모델을 만드는 것이므로 결과 해석이 어렵다는 단점을 지닌 딥러닝 기법을 모델의 선택지에서 배제하였다. 본 연구에서는 XGBoost 모델을 통해 어떠한 특성이 사기 게시글 탐지에 어떠한 영향을 미치는지 해석 가능한 결과를 도출하였기 때문에 향후 연구에서는 이 결과를 이용하여 딥러닝의 다양한 알고리즘을 적용하여 새로운 모델을 만드는 것도 가능할 것이다.

둘째, 본 연구에서는 사기 여부를 판별할 수 없는 데이터를 제외해야 했다. 연구의 엄격성을 위하여 당연한 조치이기는 하지만, 사용 가능한 데이터의 수가 줄어들 수밖에 없었다. 머신러닝 모

델은 사용하는 데이터셋이 클수록 더욱 좋은 성과를 내게 되므로 향후 관련 기관의 협조를 받아 더 많은 데이터에서 사기 판별이 가능하게 된다면 더욱 좋은 모델이 도출될 수 있을 것이다.

셋째, 본 연구는 중고거래 시장에서의 사기글 예측을 위한 모델을 구성하기 위하여 최고의 성과를 내는 특성만을 선택하여 최종 모델을 구성하였다. 향후 연구로서, 최적화된 모델과 별도로 가능한 모든 특성들을 선정하고 이들을 하나씩 혹은 그룹화 시켜 모델에 추가하며 각각의 영향력을 비교 분석함으로써 개별 특성에 대한 보다 깊은 이해를 얻는 연구 또한 의미 있을 것이다.

## 7.3 학술적, 실무적 시사점

본 연구의 학술적 시사점은 다음과 같다. 첫째, 본 연구는 C2C 중고 거래시 작성되는 게시글에서 다양한 특성을 추출할 수 있다는 것을 보여주었다. 특히 그간 C2C 환경의 거래에 대한 연구 중에 LDA 기법을 활용하여 주제를 탐색하고 그를 바탕으로 사기 게시글을 탐지한 연구는 없었다. 본 연구는 C2C 환경에서의 중고거래에서 사기 행위를 판단할 때 가격 등 명시된 단편적 정보뿐 아니라 게시글 전체를 종합하여 말하고자 하는 주제를 파악하고 주제마다 자주 사용되는 단어들을 이용하여 머신러닝 모델을 연구하였다. 이러한 점은 기존의 연구들과 가장 차별되는 점이라고 볼 수 있다.

둘째, 대부분의 C2C 거래 플랫폼들의 경우 판매자가 상품에 대한 게시글을 남기는 형태로 되어 있기 때문에 꼭 중고거래라는 한정된 영역이 아닌 다른 C2C거래 영역에서도 충분히 고려해볼 수 있는 방법이라는 점에서 더욱이 의의가 있다고 판단된다.

셋째, 기존의 거짓말과 사기에 대한 연구들은 특정 사실을 이야기할 때 거짓을 이야기하는 경우는 보다 구체적이고 상세한 근거를 들어 이야기하고, 진실을 이야기하는 경우는 보다 추상적이고

간략하게 이야기한다고 하였다. 본 연구의 결과는 일부 기존 연구 결과를 강화하지만 일부 연구들과는 대비되는데 그 이유로는 거짓 글을 작성하는 상황과 목적이 다르기 때문으로 추정된다. 즉, 본 연구는 온라인 환경에서의 거짓말과 아직 실현되지 않은 거짓말이라는 특징이 어떠한 점을 강화시키고 어떠한 점을 약화시키는지 실증분석을 통해 구체적으로 보였다. 따라서 상황과 목적에 따라서 거짓 글에서 어떠한 특징이 발현되는지 인과관계를 구체화하여 향후 관련 연구의 초석이 될 것을 기대한다.

실무적 시사점은 다음과 같다. 첫째, 본 연구는 C2C 중고 거래 분야에서 사기 이용자를 적발하는 기존의 방식이 가지는 제약에서 자유로운 모델을 만들었다는 의미가 있다. 이전까지 사기 게시글의 판단여부는 판매자의 전화번호, 계좌번호를 조회하여 이전에 사기 이력이 있는지 조회하는 방식으로 이루어졌다. 하지만 전화번호와 계좌번호는 언제든지 변경이 가능하며 타인의 전화번호와 계좌번호를 사용할 가능성도 존재한다. 즉 판매자가 마음만 먹으면 적발이 어려웠다. 또한 이러한 방법은 처음 사기를 의도하거나, 사기를 저질렀으나 적발되지 않은 경우는 알아낼 수 없다는 한계를 가지고 있다. 그러나 본 연구에서 제시한 모델은 게시글 만을 보고 사기 게시글 여부를 판별하기 때문에 이러한 제약에서 자유롭다. 즉, 사기 행위자가 이력을 바꾸거나, 사기 행위자로 적발된 이력이 없더라도 사기 의심 글을 판별해 낼 수 있다.

둘째, C2C 중고거래 플랫폼에서는 본 연구의 결과를 이용하여 판매자 인증 방법을 마련하거나 플랫폼의 신뢰성을 높이는 데 이용할 수 있다. 즉, 게시글이 올라오면 거래 전에 의심이 가는 게시글의 판매자와 연락하여 확인을 할 수 있을 것이다. 온라인 C2C 중고 거래 규모가 증가하며 모든 판매자들에 대해 이러한 방법을 사용하기는 어려우나 본 연구의 모델을 사용한다면 선별적 접근이 가능하여 이러한 방법을 시행하는데 드는 비용을 줄여 줄 수 있을 것이다. 나아가 이러한 조치 시 판매자

와 연락이 불가능한 경우 혹은 절대적으로 사기 글로 예측되는 경우에는 게시글을 읽는 이용자에게 사전에 경고를 주는 방식으로 사기를 예방할 수 있을 것으로 기대한다. 이러한 장치들이 도입되면 사기 행위자가 사기를 하기 어렵다는 것을 인지하게 되고 구매자들의 플랫폼에 대한 신뢰는 확대되어 C2C 중고거래 시장의 규모가 더욱 증가하고 안전성이 한층 높아질 것이다.

셋째, C2C형태의 거래들이 폭발적으로 성장하는 시기에 정상 게시글과 비교하여 사기 게시글이 가지는 특성을 밝혀내어 C2C 중고거래 플랫폼을 운영하는 실무자들과 사기 행위를 억제하고자 하는 관련 기관에도 도움이 될 것으로 기대한다. 또한 본 연구의 모델은 중고거래 뿐 아니라 C2C 형태의 다양한 플랫폼에도 쉽게 수정 적용될 수 있어 C2C 중고 거래 플랫폼을 넘어 더 큰 사회적 효용이 창출될 것이라 기대된다.

## 참고 문헌

- [1] 김진홍, 안희동, “코로나 불황 속 상반된 중고 시장의 흥행요인 분석”, *한국정보처리학회 학술대회논문집*, 제27권, 제2호, 2020, pp. 151-152.
- [2] 김하정, 조지영, 광영태, “블록체인을 활용한 P2P 중고거래 플랫폼”, *한국정보과학회 학술 발표논문집*, 2019, pp. 1645-1647.
- [3] 더치트, “피해 사례통계”, 2021, accessed Mar 28, 2021, Retrieved from [https://thecheat.co.kr/rb/?mod=\\_statistics](https://thecheat.co.kr/rb/?mod=_statistics)
- [4] 문옥영, *한국어 진술서에서 책임회피 시 나타나는 거짓의 언어·심리적 특징* (석사학위논문), 경기대학교 일반대학원, 수원, 2011.
- [5] 이경남, 전계형, “블록체인을 이용한 중고거래 플랫폼 개선방안 연구”, *디지털융복합연구*, 제16권, 제9호, 2018, pp. 133-145.
- [6] 이나은, 이상원, “컴퓨터 기반 매개 커뮤니케이션 내 준 언어적 신호에 대한 탐색적 연구”,

- 한국 HCI 학회 학술대회, 2017, pp. 746-749.
- [7] 이동은, “우리의 지갑을 노리는 해커, 사이버 사기의 진화”, *KISO 저널*, 제35권, 2019, pp. 52-60.
- [8] 이보한, 나중연, “소비자 간 거래 플랫폼에서의 신뢰의 구성과 형성요인”, *소비자학연구*, 제31권, 제3호, 2020, pp. 167-191.
- [9] 이석준, *거짓 진술에서 SCAN의 은폐 탐지율에 관한 연구* (석사학위논문), 경기대학교 일반대학원, 수원, 2015.
- [10] 이세정, *거짓말 행동 특징에 대한 신념: 軍수사관 중심으로* (석사학위논문), 경기대학교 행정·사회복지대학원, 경기도, 2018.
- [11] 조유빈, “시장 규모 20조, 중고장터의 ‘이유 있는’ 변신”, *시사저널*, 2020. 3. 5, Retrieved from <https://www.sisajournal.com/news/article-View.html?idxno=196345>.
- [12] 황현정, 문현수, 이영석, “온라인 중고시장에서 판매글 신뢰도 분석”, *한국정보과학회 학술 발표논문집*, 2017, pp. 1853-1855.
- [13] Akehurst, L., G. Köhnken, A. Vrij, and R. Bull, “Lay persons’ and police officers’ beliefs regarding deceptive behaviour”, *Applied Cognitive Psychology*, Vol.10, No.6, 1996, pp. 461-471.
- [14] Barse, E. L., H. Kvarnstrom, and E. Jonsson, “Synthesizing test data for fraud detection systems”, *Paper presented at the Proceedings of the 19th Annual Computer Security Applications Conference*, Las Vegas, USA, 2003.
- [15] Bond Jr, C. F. and B. M. DePaulo, “Accuracy of deception judgments”, *Personality and social psychology Review*, Vol.10, No.3, 2006, pp. 214-234.
- [16] Brown, P. F., S. A. Della Pietra, V. J. Della Pietra, J. C. Lai, and R. L. Mercer, “An estimate of an upper bound for the entropy of English”, *Computational Linguistics*, Vol.18, No.1, 1992, pp. 31-40.
- [17] Chang, J., S. Gerrish, C. Wang, J. Boyd-Graber, and D. Blei, “Reading tea leaves: How humans interpret topic models”, *Advances in Neural Information Processing Systems, Vol.22*, 2009, pp. 288-296.
- [18] Chen, T. and C. Guestrin, “Xgboost: A scalable tree boosting system”, *Paper presented at the Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining*, New York, USA, 2016.
- [19] Chua, C. E. H. and J. Wareham, “Fighting internet auction fraud: An assessment and proposal”, *Computer*, Vol.37, No.10, 2004, pp. 31-37.
- [20] de Roux, D., B. Perez, A. Moreno, M. D. P. Villamil, and C. Figueroa, “Tax fraud detection for under-reporting declarations using an unsupervised machine learning approach”, *Paper presented at the Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, New York, USA, 2018.
- [21] Dimoka, A., Y. Hong, and P. A. Pavlou, “On product uncertainty in online markets: Theory and evidence”, *MIS Quarterly*, Vol.36, No.2, 2012, pp. 395-426.
- [22] Ekman, P. and M. O’Sullivan, “Who can catch a liar?”, *American Psychologist*, Vol.46, No.9, 1991, pp. 913-920.
- [23] Fanning, K., K. O. Cogger, and R. Srivastava, “Detection of management fraud: A neural network approach”, *Intelligent Systems in Accounting, Finance and Management*, Vol.4, No.2, 1995, pp. 113-126.
- [24] Fernando, A. G., B. Sivakumaran, and L. Suganthi, “Comparison of perceived acquisition value sought by online second-hand and new goods shoppers”, *European Journal of Marketing*, Vol.52, No.7/8, 2018, pp. 1412-1438.
- [25] Ford, B. J., H. Xu, and I. Valova, “A real-time

- self-adaptive classifier for identifying suspicious bidders in online auctions”, *The Computer Journal*, Vol.56, No.5, 2013, pp. 646-663.
- [26] Ghosh, S. and D. L. Reilly, “Credit card fraud detection with a neural-network”, *Paper presented at the Proceedings of the 27th Hawaii International Conference*, Wailea, USA, 1994.
- [27] Gupta, A., “The evolution of fraud: Ethical implications in the age of large-scale data breaches and widespread artificial intelligence solutions deployment”, *International Telecommunication Union Journal*, Vol.1, 2018, pp. 1-7.
- [28] Han, H., W.-Y. Wang, and B.-H. Mao, “Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning”, *Paper presented at the Proceedings of the International Conference on Intelligent Computing*, Hefei, China, 2005.
- [29] Hart, C. L., L. P. Hudson, D. G. Fillmore, and J. D. Griffith, “Managerial beliefs about the behavioral cues of deception”, *Individual Differences Research*, Vol.4, No.3, 2006, pp. 176-183.
- [30] Jones, K. S., “A statistical interpretation of term specificity and its application in retrieval”, *Journal of Documentation*, Vol.28, No.1, 1972, pp. 11-21.
- [31] Kalman, Y. M. and D. Gergle, “CMC cues enrich lean online communication: The case of letter and punctuation mark repetitions”, *Paper presented at the Proceedings of the 5th Mediterranean Conference on Information Systems*, Tel Aviv, Israel, 2010.
- [32] KOTRA, “중국 중고시장 성장”, 2020, Retrieved from <https://news.kotra.or.kr/user/globalBbs/kotranews/782/globalBbsDataView.do?setIdx=243&dataIdx=179992>.
- [33] Lemel, R., “C2C E-commerce: The state of academic research in disposing goods online”, *Copyright 2020 by Institute for Global Business Research*, Nashville, TN, USA, 12, 2020.
- [34] Li, Y. and L. Chen, “Risk evaluation for C2C e-commerce via an improved credit counting method”, *Internet Technology Letters*, Vol.222, No.3, 2020.
- [35] Liebman, N. and D. Gergle, “It’s (Not) simply a matter of time: The relationship between CMC cues and interpersonal affinity”, *Paper presented at the Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, 2016.
- [36] Lin, J. W., M. I. Hwang, and J. D. Becker, “A fuzzy neural network for assessing the risk of fraudulent financial reporting”, *Managerial Auditing Journal*, Vol.18, No.8, 2003, pp. 657-665.
- [37] Little, B. B., W. L. Johnston Jr, A. C. Lovell, R. M. Rejesus, and S. A. Steed, “Collusion in the US crop insurance program: Applied data mining”, *Paper presented at the Proceedings of the 2002 SIAM International Conference on Data Mining*, 2002.
- [38] Luhn, H. P., “A statistical approach to mechanized encoding and searching of literary information”, *IBM Journal of Research and Development*, Vol.1, No.4, 1957, pp. 309-317.
- [39] Lundberg, S. and S.-I. Lee, “A unified approach to interpreting model predictions”, arXiv preprint arXiv:1705.07874, 2017.
- [40] Maes, S., K. Tuyls, B. Vanschoenwinkel, and B. Manderick, “Credit card fraud detection using Bayesian and neural networks”, *Paper presented at the Proceedings of the 1st International Naiso Congress on Neuro Fuzzy Technologies*, 2002.
- [41] Newman, D., J. H. Lau, K. Grieser, and T. Baldwin, “Automatic evaluation of topic coherence”, *Paper presented at the Proceedings of the 2010 Annual Conference*, 2010.

- [42] Newman, M. L., J. W. Pennebaker, D. S. Berry, and J. M. Richards, "Lying words: Predicting deception from linguistic styles", *Personality and Social Psychology Bulletin*, Vol.29, No.5, 2003, pp. 665-675.
- [43] Rasheed, L. O. and A. Olukemi, "Reputation system for fraud detection in nigerian consumer-to-consumer e-commerce", *Journal of Computer Science*, Vol.7, No.2, 2019, pp. 49-60.
- [44] Shapley, L. S., *Notes on the N-person Game--II: The Value of an N-person Game*, Rand Corporation, 1951.
- [45] Sinayobye, J. O., F. Kiwanuka, and S. K. Kyanda, "A state-of-the-art review of machine learning techniques for fraud detection research", *Paper presented at the Proceedings of the 2018 Symposium on Software Engineering in Africa*, 2018.
- [46] Smith, N., *Reading between the Lines: An Evaluation of the Scientific Content Analysis Technique (SCAN)*, London, England: Home Office, 2001.
- [47] Tsang, S., Y. S. Koh, G. Dobbie, and S. Alam, "Detecting online auction shilling frauds using supervised learning", *Expert Systems with Applications*, Vol.41, No.6, 2014, pp. 3027-3040.
- [48] Vrij, A., *Detecting Lies and Deceit: Pitfalls and Opportunities*, New Jersey, USA: John Wiley & Sons, 2008.
- [49] Vrij, A., K. Edward, and R. Bull, "People's insight into their own behaviour and speech content while lying", *British Journal of Psychology*, Vol.92, No.2, 2001, pp. 373-389.
- [50] Wang, Y. and W. Xu, "Leveraging deep learning with LDA-based text analytics to detect automobile insurance fraud", *Decision Support Systems*, Vol.105, 2018, pp. 87-95.
- [51] Xue, H.-Y. and D.-H. Yang, "Implementing circular consumption by means of second-hand goods market", *Paper presented at the Proceedings of the 2010 International Conference on Management and Service Science*, 2010.
- [52] Yamamoto, H. and H. Ohshima, "Proactive or reactive? Platform governance strategy in C2C marketplace", *Paper presented at the Proceedings of the Pacific Asia Conference on Information Systems (PACIS)*, 2017.
- [53] Yoshida, T. and H. Ohwada, "Shill bidder detection for online auctions", *Paper presented at the Proceedings of the Pacific Rim International Conference on Artificial Intelligence*, 2010.
- [54] Zainuddin, A., J. Junaidi, and R. D. Putra, "Design of e-commerce payment system at tokopedia online shopping site", *Aptisi Transactions On Management*, Vol.1, No.2, 2017, pp. 143-155.
- [55] Zuckerman, M., B. M. DePaulo, and R. Rosenthal, "Verbal and nonverbal communication of deception", In *Advances in experimental social psychology* (Vol.14, pp. 1-59). Amsterdam, Netherlands: Elsevier, 1981.
- [56] Štrumbelj, E. and I. Kononenko, "Explaining prediction models and individual predictions with feature contributions", *Knowledge and Information Systems*, Vol.41, No.3, 2014, pp. 647-665.

## **A Study on the Fraud Detection in an Online Second-hand Market by Using Topic Modeling and Machine Learning**

Dongwoo Lee\* · Jinyoung Min\*\*

### **Abstract**

As the transaction volume of the C2C second-hand market is growing, the number of frauds, which intend to earn unfair gains by sending products different from specified ones or not sending them to buyers, is also increasing. This study explores the model that can identify frauds in the online C2C second-hand market by examining the postings for transactions. For this goal, this study collected 145,536 field data from actual C2C second-hand market. Then, the model is built with the characteristics from postings such as the topic and the linguistic characteristics of the product description, and the characteristics of products, postings, sellers, and transactions. The constructed model is then trained by the machine learning algorithm XGBoost. The final analysis results show that fraudulent postings have less information, which is also less specific, fewer nouns and images, a higher ratio of the number and white space, and a shorter length than genuine postings do. Also, while the genuine postings are focused on the product information for nouns, delivery information for verbs, and actions for adjectives, the fraudulent postings did not show those characteristics. This study shows that the various features can be extracted from postings written in C2C second-hand transactions and be used to construct an effective model for frauds. The proposed model can be also considered and applied for the other C2C platforms. Overall, the model proposed in this study can be expected to have positive effects on suppressing and preventing fraudulent behavior in online C2C markets.

**Keywords:** *Keywords: C2C second-hand market, Machine Learning, Fraud Detection Model*

---

\* Researcher, RT DataLab

\*\* Corresponding Author, Associate Professor, College of Business, Chosun University

## ◎ 저 자 소 개 ◎



**이 동 우 (dong461@naver.com)**

조선대학교 경상대학 경영학부를 졸업하고, 동교 대학원에서 석사 학위를 취득했다. 현재 알티데이터랩에서 연구원으로 재직 중이다. 주요 관심 분야는 Fraud detection, Text Mining, Topic Modeling 등이다.



**민 진 영 (saharamin@chosun.ac.kr)**

현재 조선대학교 경상대학 경영학부에서 부교수로 재직 중이며 KAIST 경영대학에서 경영정보시스템 전공으로 박사학위를 취득하였다. 주요 관심 분야는 Social Media, Human and Computer Interaction, Knowledge Management 등이다. Computers in Human Behavior, International Journal of Information Management, Journal of the Association for Information Science and Technology, 경영학 연구 등에 논문을 게재한 바 있다.

논문접수일 : 2021년 05월 17일

게재확정일 : 2021년 08월 06일

1차 수정일 : 2021년 07월 13일