

SNS 텍스트 기반 마약 의심 거래 탐지 알고리즘에 대한 연구: 패턴 및 맥락 분석을 중심으로

안 상 선*

국민대학교

본 연구는 SNS 플랫폼에서 급증하는 마약 거래와 지속적으로 진화하는 은어 사용에 대응하기 위한 효과적인 탐지 방안을 제시하는 것을 목적으로 한다. 마약 거래 의심 게시물과 정상 게시물을 혼재된 데이터 이용해 자연어 처리와 머신러닝 기법을 활용한 체계적인 분석을 수행했다. 분석 결과, SNS 마약 거래 텍스트에서는 대유, 비유, 쌍관, 묘사의 네 가지 언어적 패턴이 발견되었으며, TF-IDF 분석과 Word2Vec 분석을 통해 이들 텍스트가 조밀한 의미 군집을 형성하고 해시태그를 통한 은밀한 코드화가 이루어지고 있음을 확인했다. 이상탐지 모델 실험에서는 로지스틱 회귀모델이 99.89%의 정확도를 달성했으며, LIME과 SHAP 분석을 통해 마약 거래 텍스트가 일반 상거래의 특징을 교묘하게 모방하는 패턴이 발견되었다. 본 연구는 SNS 마약 거래의 언어적 특성을 체계적으로 규명하고 효과적인 탐지 방안을 제시함으로써, 온라인 마약 거래 문제에 대한 실효성 있는 대응 방안을 제공했다는 데 의의가 있으며, 다양한 사이버 범죄 탐지에도 확장 적용이 가능할 것으로 기대된다.

주제어: 인공지능, 마약거래, 자연어분석, 사이버범죄, 머신러닝

* 주저자: 안상선/국민대학교 소프트웨어융합대학원 겸임교수, 매일경제 사외벤처 (주)M-Robo 대표 /서울시 강서구 마곡중앙로 161-17/E-mail: sangsun.ahn@m-robo.com

I. 서론

1. 연구 배경

우리나라의 2023년 마약사범은 27,611명으로 역대 최대치를 기록했으며, 전년(18,395명) 대비 50.1% 증가했다. 특히 10대 마약사범이 1,477명으로 전년 대비 207% 급증했고, 20대 마약사범은 8,368명으로 전체 마약사범의 35.6%를 차지했다.¹⁾ 전체 마약류 압수량은 998kg으로 전년 대비 24% 증가했다.

또한 마약류 유통 방식이 기존의 대면 거래에서 온라인 비대면거래 방식으로 변화하고 있어, 다크웹과 보안메신저를 이용한 신종 범죄가 증가하는 추세이다. 과거에는 ‘오른손-왼손’ 방식이라 불리는 대면거래에서, ‘던지기’ 방식이라고 하는 온라인 비대면거래로 전환됐는데, 이 같은 방식은 총책, 관리책, 드라이퍼(배달책) 등으로 이루어진 점조직 형태로 운영되며, 다크웹, 보안메신저, 가상화폐의 익명성을 적극 활용하기 때문에 검거 및 단속이 어렵다. 특히 온라인 기반 마약거래는 특히 청소년들의 접근을 용이하게 만들어 사회적 우려가 커지고 있다.

이러한 상황에서 SNS상의 마약 은어 사용은 문제를 더욱 심각하게 만들고 있다. ‘떨’(대마초), ‘케이’(케타민), ‘캔디’(엑스터시), ‘아이스’(필로폰), ‘허브’(합성대마) 등의 은어를 사용함으로써 마약에 대한 심리적 진입장벽을 낮추고 있으며, 특히 청소년들이 이를 일상적 용어로 인식하게 만들어 마약의 위험성을 약화시키고 있다.

더욱이 익명성이 보장되는 SNS 플랫폼에서는 ‘해 뜨다’(밀수 성공), ‘해 지다’(밀수 적발) 등의 은어를 활용한 조직적인 밀수와 유통이 증가하고 있으며, ‘풍당’(음료에 마약 투입)과 같은 신종 범죄수법은 성범죄나 금전 갈취 등 2차 범죄로 이어질 위험도 높이고 있다. 마약 관련 은어가 지속적으로 진

화하고 새로운 플랫폼으로 이동하면서, 법 집행기관의 효과적인 대응이 더욱 어려워지는 상황이다.

2. 연구 목적

본 연구는 급증하는 SNS 기반 마약 거래와 지속적으로 진화하는 마약 은어에 대응하기 위한 기술적 해결 방안을 제시하는 것을 목적으로 한다. 현재 마약 거래 단속에서 가장 큰 어려움은 은어의 빠른 변화와 새로운 플랫폼으로의 이동이다. 특히 ‘작대기’(필로폰이 든 주사기), ‘한잔 하자’(필로폰 투약), ‘고속도로’(상습 투약 흔적) 등과 같이 일상적인 단어를 마약 은어로 활용하는 경우가 많아, 기존의 키워드 기반 필터링으로는 탐지가 어렵다.

본 연구에서 제안하는 알고리즘은 자연어 처리 기술과 딥러닝을 결합하여 문맥을 이해하고, 새롭게 등장하는 은어까지 탐지할 수 있는 진보된 방식을 제시한다.

구체적으로 다음과 같은 방법론을 제시한다. 첫째, SNS 게시물과 해시태그, 댓글 등에서 마약 거래 관련 콘텐츠를 식별하는 이상탐지 모델을 개발한다. 이 모델은 게시물의 텍스트 패턴, 해시태그 사용 방식, 계정 간 상호작용 패턴 등을 종합적으로 분석하여 의심스러운 활동을 탐지한다. 둘째, 최신 언어 생성 모델을 활용하여 새롭게 등장하는 은어를 학습하고 예측하는 시스템을 구축한다. 이를 통해 마약 거래자들이 사용하는 새로운 은어에 선제적으로 대응할 수 있다.

본 연구의 기대효과는 다음과 같다. 첫째, 마약 거래 탐지의 정확도와 효율성이 크게 향상될 것으로 기대된다. 특히 실시간 모니터링이 가능해져, 거래가 이루어지기 전에 예방적 조치를 취할 수 있다. 둘째, 청소년들의 마약 접근을 조기에 차단할 수 있다. 2023년 10대 마약사범이 전년 대비 207% 증가

1) 2023년 마약류 범죄백서

한 현 상황에서, 이는 매우 중요한 의미를 갖는다. 셋째, 수사기관의 인력과 자원을 더욱 효율적으로 활용할 수 있게 된다. 알고리즘이 1차적으로 의심 게시물을 선별함으로써, 수사관들은 보다 전문적이고 심층적인 수사에 집중할 수 있다.

더불어 본 연구는 마약 범죄 외에도 다른 형태의 사이버 범죄 탐지에도 활용될 수 있는 방법론을 제시했다. 특히 성매매, 불법 도박, 보이스피싱 등 은어를 활용한 다양한 범죄 유형의 SNS 활동 탐지에도 적용이 가능하다. 이는 궁극적으로 우리 사회의 사이버 범죄 대응 역량을 전반적으로 향상시키는 데 기여할 것으로 기대된다.

II. 본론

1. SNS 마약 거래 포스팅의 언어적 특성

은어는 ‘특정 부류의 사람들이 비밀, 유희, 친교, 성별 등의 목적으로 공용어에는 존재하지 않는 새로운 표현들을 만들어 사용함으로써 일반 대중이 쉽게 알아듣지 못하고 은밀함을 느끼게 하는 특수어?’로 정의된다.

은어는 사용 맥락과 목적에 따라 다양하게 분류된다. 연구자들의 분류 체계를 보면, 김종훈, 박영섭(2005)은 대학생, 범죄인, 걸인의 은어로, 김홍석(2011)은 궁중어, 범죄인, 심마니의 은어로 구분했다. 범죄 은어는 범죄 행위를 은폐하기 위한 독특한 특성을 보인다. 범죄의 다양화와 빈도 증가에 따라 은어도 빠르게 진화하며, 특히 마약 범죄 분야에서 이러한 경향이 두드러진다.

마약 거래에서 사용되는 은어의 핵심적 특성은 은폐성(隱蔽性)에 있다. 이는 특정 집단 내부에서만

의미가 공유되고 외부로는 그 의미가 은폐되는 특성을 의미한다.³⁾ 은어의 은폐성은 단일적인 성질이 아닌 복합적인 성질을 가지고 있으며, 이는 다른 특수어의 특성들과 결합하여 나타난다.

예를 들어, 유행어의 핵심적 자질인 편리성과 일시성, 비속어의 핵심적 자질인 신기성과 저속성이 동시에 드러나면서 은폐성을 형성한다. 마약 거래 은어의 경우, 이러한 은폐성은 불법 거래의 은닉과 수사기관의 추적 회피라는 실질적 목적을 가진다.

랜덤채팅 앱 속 ‘마약은어’



The JoongAng

[그림 1] 채팅앱에서의 마약 거래 은어⁴⁾



[그림 2] SNS에서의 마약 거래 은어⁵⁾

2) 김종훈, 박영섭 외 2명(2005), 『은어·비속어·직업어』, 집문당, p.13-15.
 3) 안정근 외 2명(2013), 「은어사용의 사회언어학적 고찰-세대별, 성별 차이를 중심으로」, 『건지인문학』 제9집, p.285
 4) “아이스 하실 분 10대 마약·성매매 채팅...정부 알고도 놔둔다 왜”. 2024.10.29. 중앙일보
 5) “아이스작대기 검색하니 마약판매글 주르륵...삭제엔 35일 걸려”. 2022.10.09. 이데일리

최근 소셜 네트워크 서비스(SNS) 플랫폼의 급속한 발달은 마약 거래에서 사용되는 은어의 진화 양상에 큰 영향을 미치고 있다. 특히 주목할 만한 특징으로는 일상 용어의 은어화, 다중 의미층 구조의 발달, 그리고 빠른 변화 주기를 들 수 있다.

먼저, 일상 용어의 은어화 현상이 두드러지게 나타나고 있다. 예컨대 ‘작대기’는 주사기를, ‘한 잔’은 1회 투약량을, ‘고속도로’는 상습 투약 흔적을 지칭하는 등 일상적인 단어들에 마약 거래 용어로 전용되고 있다. 이러한 방식은 일반적인 SNS 대화와의 구분을 모호하게 만들어 단속을 회피하는 수단으로 활용되고 있다.⁶⁾

또한, 은어의 의미 구조가 다층화되는 특징도 관찰된다. 표면적 의미와 실제 의미가 여러 층위로 구성되어 있는데, 대표적인 예로 ‘해 뜨다’와 ‘해 지다’라는 자연 현상을 각각 밀수 성공과 적발이라는 범죄 행위와 연관시키는 은유적 표현이 사용되고 있다.⁷⁾

마지막으로, 온라인 플랫폼의 특성을 활용한 빠른 변화 주기가 특징적이다. 수사기관의 탐지를 피하기 위해 지속적으로 새로운 은어가 생성되며, 이러한 은어들은 SNS 플랫폼의 즉시성과 확산성을 통해 신속하게 전파되고 변형되는 양상을 보인다.

2. 유해물 탐지를 위한 연구 방법론

(1) 머신러닝 알고리즘을 이용한 탐지

소셜 네트워크 서비스(SNS)에서의 마약 거래 포스팅을 효과적으로 탐지하기 위해, 본 연구에서는 전통적인 머신러닝 기법과 첨단 자연어 처리 기술을 결합한 다층적 접근 방식을 채택했다.

이러한 접근 방식은 Christopher J. Rhodes의 연

구에서 제시된 개방형 정보원(OSINT)을 활용한 은밀한 소셜 네트워크 구축 기법과 맥을 같이 한다. 이 연구에서는 공개된 정보를 통해 숨겨진 연결망을 파악하는 방법을 제시했는데, 이는 마약 거래 네트워크를 식별하는 데 유용하게 적용될 수 있다.

또한, Neri와 Pettoni가 개발한 다국어 텍스트 마이닝 검색 엔진 ‘Stalker’의 개념을 응용하여, 일상어 외 은어와 이모티콘 등 다양한 기호로 구성된 마약 관련 포스팅을 효과적으로 탐지하고 분석할 수 있는 방법을 고려했다.

특히, 텍스트 데이터의 특성을 고려한 자연어 전처리 기술과 다양한 머신러닝 알고리즘을 통한 이상탐지, 그리고 최신 언어 모델을 활용한 패턴 학습을 통해 SNS상에서 발생하는 불법 마약 거래 관련 게시물을 효과적으로 식별하고 차단하는 방법을 고려했다.

먼저 텍스트 데이터에 대한 철저한 전처리 과정을 거친다. 형태소 분석을 통해 텍스트를 의미 있는 최소 단위로 분해하고, 불용어 제거를 통해 분석에 불필요한 요소를 제거하며, 정규화 과정을 통해 텍스트의 일관성을 확보한다. 이러한 전처리된 데이터를 바탕으로, 세 가지 주요 머신러닝 알고리즘을 적용하여 일차적인 이상탐지를 수행한다.

로지스틱 회귀 분석은 텍스트의 특성을 바탕으로 마약 거래 게시물일 확률을 예측한다. 이 방법은 각 특성의 가중치를 명확하게 파악할 수 있어, 어떤 단어나 패턴이 마약 거래 탐지에 주요한 영향을 미치는지 파악할 수 있다는 장점이 있다. 실험 결과, 특히 은어와 관련된 특성들이 높은 가중치를 가지는 것으로 나타났으며, 이는 마약 거래 탐지에 있어 은어 탐지의 중요성을 입증한다.

K-최근접 이웃 알고리즘은 텍스트의 벡터 공간에서 유사한 게시물들을 군집화하여 분석한다. 본

6) “마약매매까지... 코인정보방 불법 온상 방치”, 2022.05.02. 전자신문.

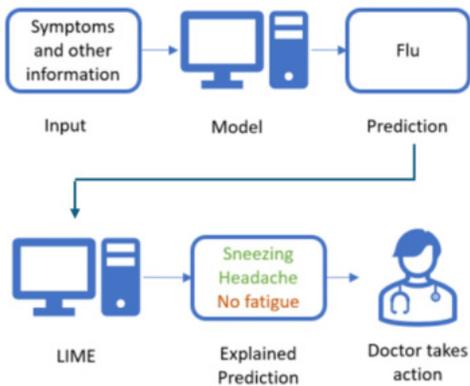
7) 최태석(2022). “한·중 마약 은어 비교 분석”. 한국외국어대학교 석사학위논문

연구에서는 코사인 유사도를 기반으로 한 거리 측정법을 사용하였으며, 최적의 K값은 교차 검증을 통해 결정되었다. 이 방법은 새로운 유형의 마약 거래 패턴이 기존 패턴과 유사성을 가질 경우 효과적으로 탐지할 수 있다는 장점을 보였다.

의사결정나무 알고리즘은 게시물의 특성을 계층적으로 분석하여 마약 거래 여부를 판단한다. 이 방법은 특히 규칙 기반의 명확한 의사결정 과정을 제공하여, 탐지 결과에 대한 해석이 용이하다는 장점이 있다. 실험에서는 가지치기(pruning)를 통해 과적합을 방지하고, 정보이득(information gain)을 기준으로 최적의 분할을 결정한다.

(2) 의미구조 파악을 위한 설명 가능한 AI모델

위의 마약거래 탐지를 위한 머신러닝 알고리즘은 마약거래 관련 글을 효과적으로 찾을 수 있지만, 각 단어 간의 맥락적 관계를 설명하기는 어렵다. 이를 위해 단어 간의 관계성과 숨겨진 의미 구조를 파악하기 위해 토픽모델링을 수행하였다.



[그림 3] LIME 모델⁸⁾

LDA(Latent Dirichlet Allocation) 기법을 활용하여 마약 거래 관련 게시물에서 자주 등장하는 주제들을 추출하였으며, 이를 통해 마약 거래와 관련된

언어와 표현의 군집을 파악할 수 있었다. 특히, 시간에 따른 토픽의 변화를 분석함으로써, 마약 거래 언어의 진화 패턴을 추적할 수 있었다.

LIME(Local Interpretable Model-agnostic Explanations) 알고리즘은 블랙박스 모델을 지역적으로 근사하여 예측 결과에 대한 설명 가능성을 제공하는 방법이다. 독감 예측 모델의 사례를 통해 살펴보면, LIME은 각 특성이 예측에 미치는 영향을 지역적으로 해석할 수 있게 한다. 구체적으로, 환자의 증상 중 재채기와 두통은 독감 예측에 긍정적으로 기여한 반면, 피로 증상이 없는 것은 독감 예측에 기여하지 않은 것으로 분석되었다. 이러한 LIME의 특성을 활용하여 텍스트 분류 모델에서 마약거래의 핵심 단어들을 분석한 결과, 각 모델별로 예측에 유의미하게 기여하는 단어들과 기여도가 낮은 단어들을 효과적으로 구분해 낼 수 있다.

이 같은 방법은 모델의 신뢰성 검증과 개선에 중요한 통찰을 제공할 수 있다. 즉, LIME 분석을 통해 단일 단어보다는 특정 단어들의 조합과 배열이 마약 거래 언어를 구성하는 핵심 요소임을 확인할 수 있다. 예를 들어, 일상적인 단어들도 특정 순서나 조합으로 사용될 때 은밀한 거래 의도를 내포할 수 있는지를 검증할 수 있다.

본 연구에서는 LIME 이외에 SHAP(SHapley Additive explanations)는 모델을 사용했는데, LIME이 지역적 근사를 통해 설명을 제공하는 반면, SHAP은 Local과 Global 설명을 모두 제공할 수 있다는 장점이 있다.

(3) 마약 거래 등 유해물 판별 주요 선행 연구

유해물 판별 연구는 크게 수작업 분류와 인공지능을 활용한 분류로 나눌 수 있다. 정규철 외(2004)는 텍스트 데이터에서 유해 단어의 중요도를 분석

8) “Explainable AI with LIME Library”. 2021.11.12. Emi, Medium.

하여 유해성을 구별하였고, 한병우·윤지원(2016)은 전문가가 구축한 사전을 기반으로 검색을 통해 사이트를 분석·수집하는 방법을 제시하였다. 수작업 분류의 경우, 서울시의 ‘인터넷 시민 감시단’ 운영이나 방송통신심의위원회의 디지털 성범죄 전담 부서 운영 등이 대표적이다. 이들은 24시간 이내 불법 촬영물을 삭제하는 등 적극적인 모니터링을 수행하고 있다.

인공지능을 활용한 분류는 크게 세 가지 방식으로 구분된다. 첫째, 내부 콘텐츠 분석 방식이다. 조동욱·김지영(2004)은 음향데이터의 신호 일치율을 분석해 유해성을 판별하였고, 민순호(2017)는 동영상 프레임을 추출하여 특정 값들을 정규화하는 방법을 제시하였다. 김세민 외(2010)는 피부와 인체영역을 검출하여 유해 이미지를 판단하는 방법을, 송철환·유성준(2006)은 SVM의 이미지 특징 히스토그램을 적용해 색상요소를 판별하는 방법을 연구하였다.

둘째, 신정훈(2013)이 제시한 URL 연결관계를 이용한 분석 방식이다. 이는 유해 사이트들 간의 상호 연결 링크 특성을 활용하여 빠른 판별이 가능하다는 장점이 있다.

셋째, 김병진·이상준(2019)이 연구한 메타데이터를 활용한 화이트리스트 구축 방식이다. 이는 기계학습 기반의 예측 알고리즘을 통해 유해사이트를 분류한다.

최근에는 SNS상에서의 유해물 판별 연구도 활발히 진행되고 있다. 장정현 외(2017)는 자연어처리 없이 해시태그만을 사용하여 유해정보를 판단·수집하는 방법을 제시하였다. 윤현식, 윤영호, 박현재(2020)의 연구에서는 트위터의 성매매 홍보 트윗을 대상으로 랜덤 포레스트 기법을 적용하여 0.996의 정밀도와 0.998의 AUC 값을 도출하였다. 이는 계정의 행태적 특성과 트윗 정보만으로도 효과적인 판별이 가능함을 보여준다.

그러나 이러한 기술적 발전에도 불구하고, 현재

의 유해물 판별 시스템은 여전히 한계점을 가지고 있다. 특히 성매매 알선이나 불법 촬영물 등에서 은어와 일상어가 혼용되는 최근의 경향은 기존의 판별 시스템의 효과성을 저하시키는 요인이 되고 있다.

또한 정부와 기업의 한정된 가용자원으로는 급증하는 유해물에 대한 즉각적인 대응이 어려운 실정이다. 따라서 은어와 일상어의 변화를 반영할 수 있는 새로운 판별 기술의 개발과 함께, 효율적인 자원 활용을 위한 시스템 개선이 필요한 상황이다.

III. 모델 구성 및 분석 결과

1. 데이터 수집 및 특성 분석

본 연구에서 사용된 데이터는 트위터(현재 X) 플랫폼에서 크롤링을 통해 수집된 게시물들로, 마약 거래 의심 텍스트와 일반 텍스트를 동일한 비율로 포함하고 있다.

데이터 세트는 실제 SNS상의 마약 거래 관련 은어 사용 패턴을 분석하기 위해 체계적으로 수집 및 구성되었다. 다크웹이나 메신저 등 다른 플랫폼의 데이터는 현실적인 수집의 제약으로 인해 본 연구에서는 제외되었다.

no	text
1	대략죽에서 아이스아메리카노 마시면서 시원하게 즐기는 중! #아아 #죽제 #얼음
2	더운 여름, 아이스커피가 최고의 선택이죠! 시원한 얼음아로 더워 탈출~ #아이스크...
3	우리 가게 얼음은 위생적인 지하수로 만든 보석얼음만 사용합니다! #얼음 #위생 #...
4	집에서도 편하게 아이스 하이볼 즐기세요! 잔얼을 택배배송으로 간편하게~ #하이볼 ...
5	업소용 얼음이 필요하다면? 1kg, 3kg 단위로 구매 가능합니다! #업소용 #얼... ...
3116	며칠째 품질이었던 🍷 주종 맥상 2g 10g 남음 합리적 가격 가격은 디엠으로 안내...
3117	찾고 계신가요? 풀 맥상 드라피 필리터 보장 이번 기회를 잡으세요 🍷 #풀 #맥상
3118	여기가 제일 저렴한 🍷 사탕 두통 물리맘 진짜 좋은 거 문의 꼭 주중이에요 🍷! #사탕...
3119	시흥 정양동 🍷 주종 1g 특가 🍷 안전거래 가능! 당장 연락주세요! #주종 #1g 특가
3120	이거 구하기 어려운데 🍷 고양이마취제 사기 안전거래 가능 고민하시면 늦어요! #고양...

	word	type
"아이시아메리카노", "내학축제", "축제", "얼음", "아아"		0
"아이스커피", "얼죽아", "얼음"		0
"얼음", "위생", "지하수", "보석얼음"		0
"하이볼", "얼음", "잔얼음", "택배배송", "아이스"		0
"업소용", "얼음", "1kg", "3kg"		0
...
주중,액상,2g 10g 남음		1
플,액상,드라피		1
사탕,두통,롤리팝		1
시홍,주중,1g 특가		1
고양이마취제,사키		1

[그림 4] 수집된 SNS 게시물 데이터

마약 거래 관련 텍스트의 주요 특징은 일상적인 단어를 은어로 활용하는 패턴을 보인다는 점이다. 예를 들어 “아이스”나 “떨”과 같은 단어들이 마약이나 불법 의약품을 지칭하는 데 사용되지만, 이러한 단어들은 동시에 일상적인 맥락에서도 빈번하게 사용되는 특징이 있다.

<표 1> 데이터 세트

변수	주요 특징	비고
TEXT	트위터(X)의 본문 글	한국어, 영어, 특수기호, 이모티콘 포함
WORD	트위터(X)의 주요 키워드	
TYPE	정상거래는 0 마약거래는 1	3,120개

데이터셋의 균형성을 확보하기 위해 마약 거래 의심 게시물 1,560개와 동일한 단어를 포함하지만, 일상적인 맥락에서 작성된 정상 게시물 1,560개를 수집하여 총 3,120개의 표본을 구성했다. 예를 들어, “아이스”라는 단어가 마약을 지칭하는 용도로 사용된 게시물과 “아이스커피”, “아이스크림” 등 일상적인 맥락에서 사용된 게시물을 균형있게 포함시켰다.

데이터의 현실성을 높이기 위해 게시물의 작성 시간, 해시태그 사용 패턴, 이모지 활용, 문장 구조 등 SNS 텍스트의 다양한 언어적 특성을 포함시켰다. 또한 텍스트의 길이, 특정 단어의 출현 빈도, 문맥적 연관성 등의 정량적 지표를 계산하여 실제 SNS 게시물의 특성을 최대한 반영했다.

수집된 SNS 게시물에서 나타나는 단어들의 특성을 유형별로 분석한 결과, 최태석(2022).의 연구와 같이 대유, 비유, 쌍관, 묘사로 구분되는 언어적 패턴이 발견되었다.

<표 2> 마약관련 은어 목록 및 주요특징

변수	주요 은어
대유	러쉬, E, ㄷㄷ, ㄸㄸ, 텔레, 엘시드
비유	크리스탈, 크랭크, 눈꽃, 차가운 술, 똥술, 밀가루, 얼음, 가루, 캔디, 단풍, 담배, 피로회복제, 영양제, 고속도로, 호남선, 풀, 죽다, 병원놀이, 지게꾼, 술, 한잔하다, 물건, 가루쟁이, 약장사, 허브, 교수, 곰, 산타, 총장, 마린보이, 멧덩구리, 야당, 빨대, 고기, 해지다, 사고났다, 얼음공주, 달리다, 뽕밭이, 환자, 공격을 쓰다, 개살구, 상선, 하선, 떨 보러 가자, 아이스작대기
쌍관	떨, 쨌
묘사	몰래뽕, 도리, 도리도리, 홍콩갔다, 뚜껑이 열리다, 꼭지갔다, 쭈라, 똥마이잡다, 공포의 백색가루, 백색의 유혹

먼저 대유 관계를 보이는 단어들은 ‘크리스탈’, ‘프로포폴’, ‘도통’, ‘하키퍼’ 등과 같이 실제 대상을 직접적으로 지칭하지 않고 관련된 다른 표현으로 대체하는 특징을 보였다. 특히 이러한 단어들은 일상적인 물건이나 상황을 지칭하는 단어를 차용하여 불법 물품을 암시하는 방식으로 사용되었다.

비유적 표현의 경우 ‘얼음’, ‘가루’, ‘떨다’, ‘날릴’, ‘땀배’와 같이 마약류의 형태나 사용 방식을 일상적인 행위나 상태에 빗대어 표현하는 경향이 나타났다. 이러한 비유적 표현들은 실제 대상의 특성을 간

접적으로 묘사하면서도 일반적인 대화에서도 자연스럽게 사용될 수 있는 모호성을 가지고 있다.

쌍관의 경우 ‘빙’, ‘필’과 같이 중의적 의미를 가진 단어들도 주로 사용되었는데, 이는 문맥에 따라 일상적 의미와 불법적 의미를 동시에 내포할 수 있는 특징을 보였다. 이러한 쌍관어의 사용은 단속을 피하면서도 거래 당사자들 간의 효과적인 의사소통을 가능하게 하는 전략으로 활용되었다.

묘사적 표현으로는 ‘본래봉’, ‘도리’, ‘도리도리’, ‘홍콩오더’ 등 거래 대상이나 방식을 특정한 은어로 직접 지칭하는 패턴이 발견되었다. 이러한 묘사어들은 일반적인 소통에서는 잘 사용되지 않는 특수한 용어들로, 불법 거래의 고유한 언어 체계를 형성하고 있다.

특히 주목할 만한 점은 이러한 언어적 패턴들이 단독으로 사용되기보다는 서로 유기적으로 결합되어 복합적인 의미를 형성한다는 것이다.

예를 들어, 대유적 표현과 비유적 표현이 함께 사용되거나, 쌍관어가 특정한 묘사어와 결합하여 거래의 구체적인 조건을 암시하는 등의 패턴이 관찰되었다. 이러한 다층적인 언어 사용 패턴은 불법 거래 텍스트를 식별하는 데 중요한 특징으로 활용될 수 있다.

2. 단어 사전 구축 및 전처리

수집된 SNS 텍스트 데이터의 효과적인 분석을 위해 우선 한국어 자연어 처리를 위한 말뭉치를 구축했다. 이 과정에서 KoNLPy의 Mecab 형태소 분석기를 활용하여 텍스트를 형태소 단위로 분리하고, 각 형태소의 품사를 태깅했다.

특히 마약 거래 관련 은어들의 특수성을 고려하여 사용자 정의 사전을 추가로 구축했으며, 이를 통해 은어의 형태소 분석 정확도를 향상시켰다. 전처리 과정에서는 불필요한 특수문자, 이모티콘, URL 등을 제거하고, 띄어쓰기와 문장 부호를 정규화하여

텍스트 데이터의 일관성을 확보했다.

아래 [그림 5]에서 type는 마약관련 포스팅 여부를 나타내는 변수로, 0이면 정상, 1이면 마약거래 트윗글이다. “#10g”, “#1g” 등과 같은 마약 판매와 관련 있는 문구는 마약거래 트윗글에서 등장하는 것을 볼 수 있다.

	type	!	#	#10g	#1g	#1g당	#1g부터
0	0	1	0	0	0	0	0
1	0	1	0	0	0	0	0
2	0	1	0	0	0	0	0
3	0	1	0	0	0	0	0
4	0	1	0	0	0	0	0
...
2555	0	1	0	0	0	0	0

	type	!	#	#10g	#1g	#1g당	#1g부터
203	1	1	0	0	0	0	0
204	1	1	0	0	0	0	0
205	1	1	0	0	0	0	0
206	1	1	0	0	0	0	0
207	1	1	0	0	0	0	0
...
3115	1	0	0	0	0	0	0

[그림 5] 구축된 단어 사전

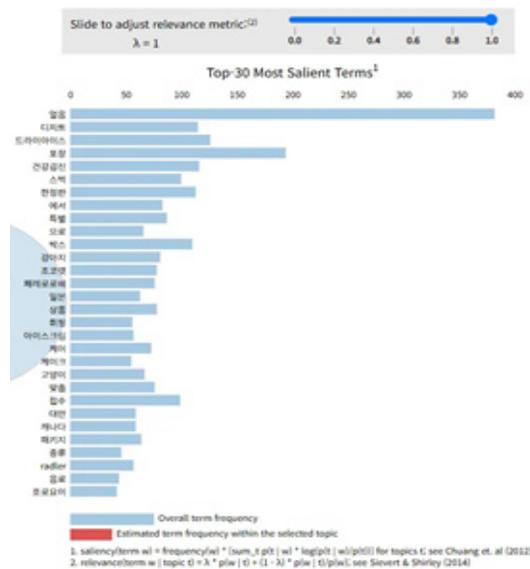
3. 빈도분석 및 워드 클라우드 시각화

전처리된 데이터를 바탕으로 단어 빈도 분석을 수행했다. 워드 클라우드(Word Cloud)는 텍스트 데이터에서 가장 빈번하게 등장하는 단어들을 시각적으로 표현하는 기법으로, 단어의 출현 빈도에 따라 글자 크기를 다르게 표시하여 직관적인 이해를 돕는다. 이를 통해 마약 거래 관련 텍스트에서 자주 사용되는 은어들의 분포와 패턴을 시각적으로 파악할 수 있었다.

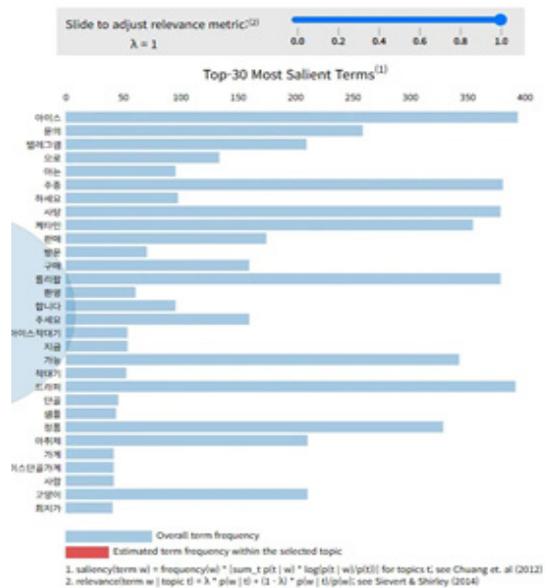
특히 마약 거래 의심 게시물에서 높은 TF-IDF 점수를 보이는 단어들은 일상적인 맥락에서는 잘 사용되지 않는 특수한 용어들이 많았으며, 이는 이러한 단어들이 해당 문서 군에서 특징적으로 사용되는 은어일 가능성을 시사한다. 이러한 TF-IDF 분석 결과는 마약 거래 관련 텍스트가 일반 텍스트와는 구별되는 독특한 어휘 사용 패턴을 가지고 있음을 정량적으로 보여준다. 또한 정상 게시물의 TF-IDF 점수가 전반적으로 더 낮은 분포를 보이는 것은, 이러한 단어들이 더 광범위한 맥락에서 일반적으로 사용되고 있음을 의미한다.

5. 토픽 모델링

LDA(Latent Dirichlet Allocation) 기반의 토픽모델링을 수행하여 텍스트 데이터에 잠재된 주제들을 추출했다. 토픽모델링은 문서 집합에서 추상적인 주제를 발견하는 통계적 기법으로, 각 문서가 어떤 주제들의 혼합으로 구성되어 있는지를 확률적으로 추정한다. 이를 통해 마약 거래 관련 텍스트에서 나타나는 주요 패턴과 의미적 구조를 파악할 수 있었다.



[그림 10] LDA 결과: 정상 게시물

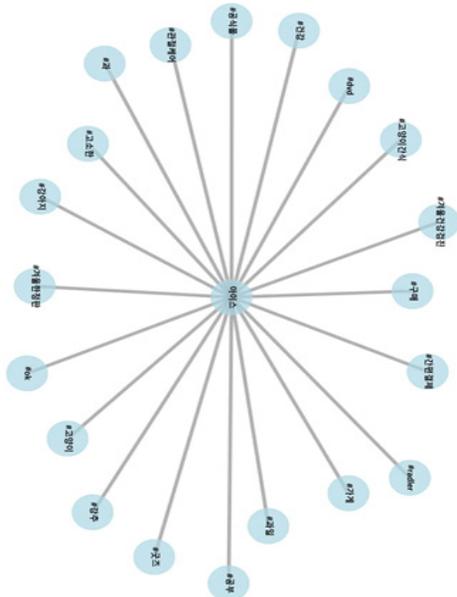


[그림 11] LDA 결과: 마약거래 게시물

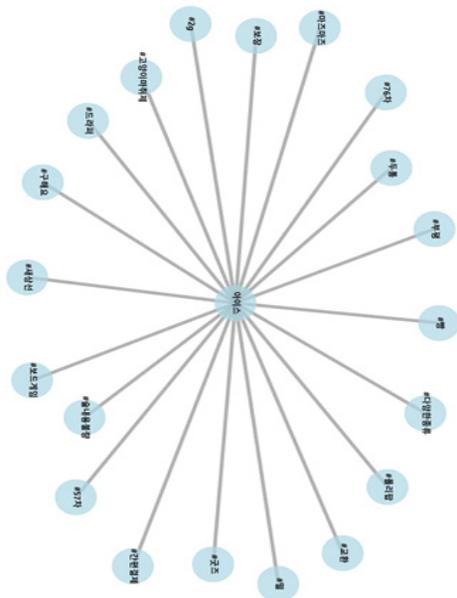
마약 게시물 및 정상 게시물의 LDA 분석 결과인 [그림 10]과 [그림 11]의 분석 결과를 비교해 보면, 정상 게시물과 마약 거래 의심 게시물 간의 주제적 특성 차이를 명확히 보여준다. 정상 게시물에서는 ‘이름’, ‘디자인’, ‘드라이아이즈’ 등 일상적인 상거래나 제품 관련 용어들이 주요 토픽을 구성하는 반면, 마약 거래 의심 게시물에서는 ‘아이스’, ‘물리’, ‘물리파람’, ‘마약제’ 등 불법 거래와 관련된 은어들이 높은 빈도로 나타났다.

Intertopic Distance Map을 살펴보면, 두 문서 군집이 명확히 구분되는 토픽 분포를 보이고 있으며, 이는 두 유형의 게시물이 서로 다른 주제적 특성을 가지고 있음을 의미한다.

Top-30 Most Salient Terms 분석에서는 마약 거래 의심 게시물의 경우 특정 은어들의 출현 빈도가 정상 게시물에 비해 현저히 높게 나타났으며, 이는 해당 단어들이 불법 거래를 위한 특수한 용어로 사용되고 있음을 시사한다. 또한 정상 게시물과 마약 거래 의심 게시물 각각의 토픽들이 서로 다른 단어 분포를 보이는 것은, 두 유형의 게시물이 서로



[그림 14] 네트워크 분석 결과: 정상 게시물



[그림 15] 네트워크 분석 결과: 마약거래 게시물

래 1,560개)을 대상으로 실험한 결과, 모델별 성능을 비교 분석한 결과, 로지스틱 회귀는 99.89%의 정확도로 가장 우수한 성능을 보였다.

<표 4> 이상탐지 모델의 성능 평가 결과

모델	Accuracy	Precision	Recall	F1 Score
Logistic Regression	99.89%	100.0%	99.79%	99.89%
K-Nearest Neighbors	99.68%	99.36%	100%	99.68%
Decision Tree	98.61%	98.72%	98.5%	98.61%

특히 정밀도 100%를 달성하여 정상 게시물을 마약 거래로 오분류하지 않았으며, 텍스트의 선형적 특성과 마약 은어의 출현 패턴을 효과적으로 학습했음을 보여준다.

K-최근접 이웃 모델은 99.68%의 정확도와 함께 100%의 재현율을 기록했다. 이는 실제 마약 거래를 완벽하게 탐지했으며, 유사한 문맥과 단어 패턴을 가진 게시물들이 효과적으로 군집화되었음을 의미한다.

의사결정나무 모델은 98.61%의 정확도를 보여 세 모델 중 가장 낮은 성능을 기록했으나, 여전히 높은 수준의 정확도를 유지했다. 이는 특정 키워드나 문구의 유무에 따른 명확한 분류 규칙이 잘 생성되었음을 나타낸다.

이러한 높은 성능의 공통적 요인은 마약 거래 은어의 높은 특이성, 게시물의 뚜렷한 문맥적 패턴, 그리고 정상과 마약 거래 게시물이 1:1로 구성된 균형 잡힌 데이터셋에 있다. 이는 텍스트 마이닝 기반의 마약 거래 탐지가 실제 현장에서도 효과적으로 적용될 수 있음을 시사한다.

8. 마약거래 이상탐지 알고리즘

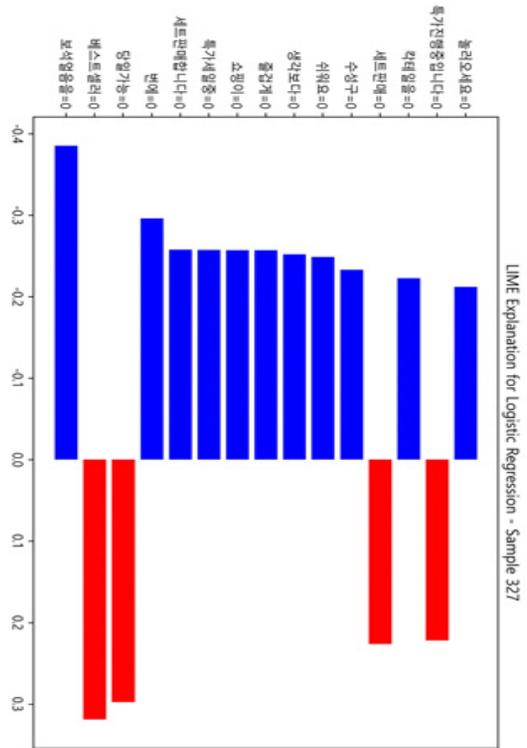
총 3,120개의 SNS 포스팅(정상 1,560개, 마약거

9. LIME 모델을 활용한 분석

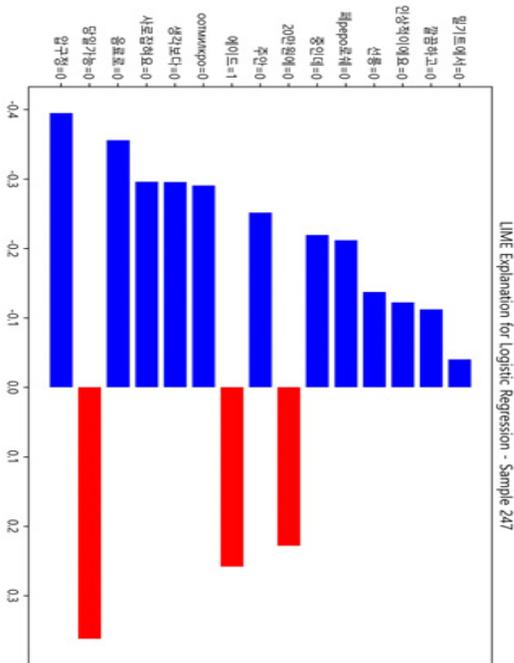
LIME 모델은 기계학습 모델의 예측을 해석하는 도구이다. 각 막대는 특정 단어나 특성이 모델의 예측에 미치는 영향력을 나타내며, 파란색은 정상글(클래스 0)로 분류되는데 기여하는 특성을, 빨간색은 마약거래(클래스 1)로 분류되는데 기여하는 특성을 보여준다.

LIME 그래프의 Y축은 각 단어나 구문이 텍스트에 존재하는지 여부(=0 또는 =1)를 나타낸다. 예를 들어, Sample 247에서 “당일가능=0”은 해당 텍스트에서 “당일가능”이라는 단어가 존재하지 않음을 의미한다. X축의 값은 각 단어가 분류에 미치는 영향력의 크기와 방향을 보여주며, 음수(파란색)는 정상 게시글로, 양수(빨간색)는 마약거래 게시글로 분류하는데 기여하는 정도를 나타낸다.

두 샘플 모두 일상적인 거래 용어들(예: “베스트셀러=0”, “세트판매=0”)이 존재하지 않음에도 이러한 특징들이 마약거래 분류에 영향을 미치는 것으



[그림 17] LIME 모델에 따른 해석: 샘플 327번



[그림 16] LIME 모델에 따른 해석: 샘플 247번

로 나타났다. 이는 마약거래 게시글들이 특정 단어의 부재를 통해서도 패턴을 형성하고 있음을 보여준다.

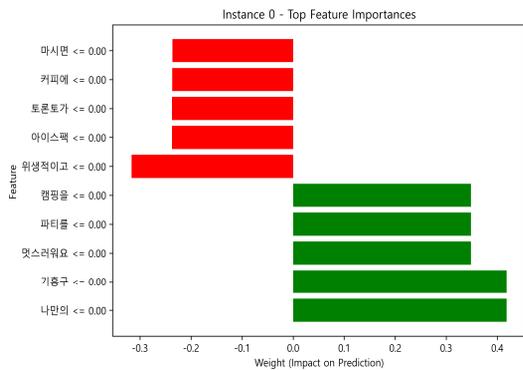
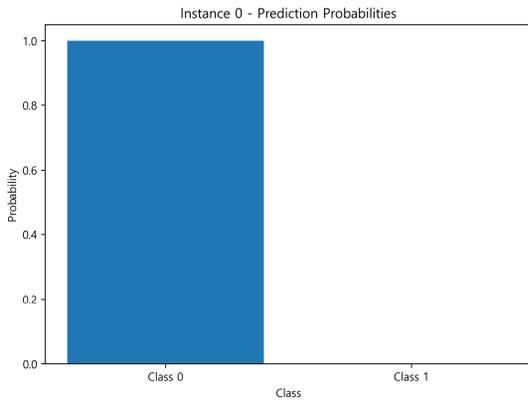
제시된 그래프들을 분석해 보면, 마약 거래 게시글들이 교묘하게 정상적인 단어들을 활용하고 있음이 드러난다. Sample 247의 경우 “당일 가능”, “에이드”, “20만 원” 등의 일상적인 단어들이 마약거래 판별에 양의 영향(빨간색)을 미치고 있다. Sample 327에서도 “베스트셀러”, “세트판매”, “특가 진행 중입니다” 같은 일반 상거래에서 흔히 사용되는 표현들이 마약거래 식별에 강한 영향을 주고 있다.

정리하면 LIME 모델을 활용한 분석을 통해서 마약거래 게시글들이 의도적으로 일반 상거래의 특징을 모방하여 detection을 회피하려는 패턴을 보여주며, LIME 분석을 통해 이러한 위장 전략이 구체적으로 어떤 단어 선택을 통해 이루어지는지 파악할 수 있다.

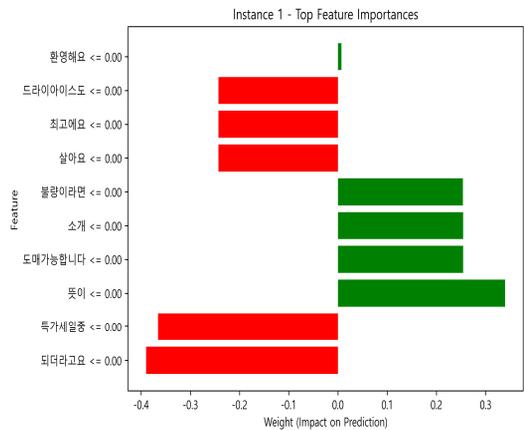
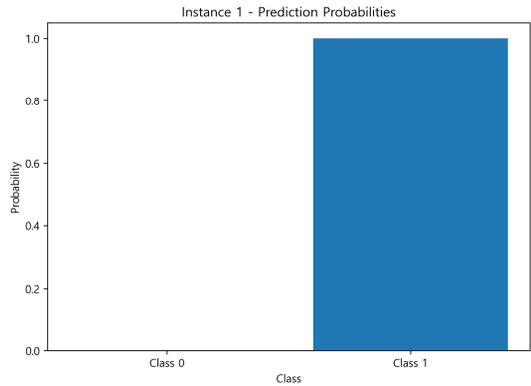
10. SHAP 모델을 이용한 설명

SHAP(SHapley Additive explanations)는 게임 이론의 Shapley 값을 기반으로 하는 머신러닝 모델 해석 기법이다. 이는 각 특성이 모델의 예측에 미치는 영향을 수학적으로 정확하게 계산하여 설명하는 방법론이다.⁹⁾

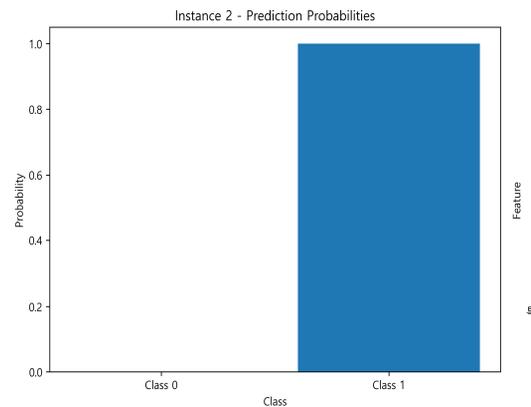
앞에서 다룬 LIME이 지역적 근사를 통해 설명을 제공하는 반면, SHAP는 게임 이론적 접근을 통해서 더 정확한 기여도를 계산한다. 즉 Local과 Global 설명을 모두 제공할 수 있다는 점이 특징인데, 이는 모델의 전체적인 동작과 개별 예측에 대한 설명이 일관성을 유지하며, 가법성을 보장하여 각 특성의 기여도 합이 전체 예측값과 일치한다는 장점이 있다.



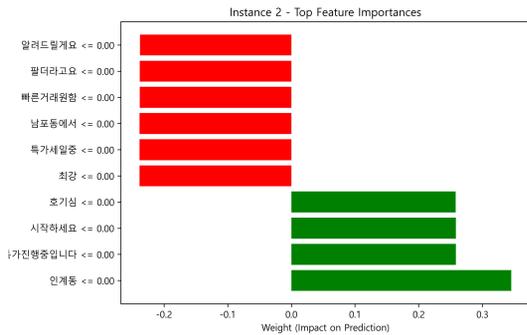
[그림 18] SHAP 모델에 따른 해석: 정상거래 글



[그림 19] SHAP 모델에 따른 해석: 마약 거래 글



9) <https://christophm.github.io/interpretable-ml-book/shap.html#fnref44>



[그림 20] SHAP 모델에 따른 해석: 다른 마약거래 글

위 [그림 18], [그림 19], [그림 20]에서 SHAP 모델을 이용한 분석은 세 가지 정보를 제공한다. 첫째, 왼쪽의 예측 확률 그래프는 각 인스턴스가 정상 게시글(Class 0)과 마약 거래 게시글(Class 1)로 분류될 확률을 보여준다. 둘째, 오른쪽의 특성 중요도 그래프는 각 단어가 분류에 미치는 영향력을 녹색(정상 게시글)과 빨간색(마약 거래 게시글)으로 시각화한다. 셋째, 각 인스턴스별로 서로 다른 특성들의 영향력을 개별적으로 분석하여 지역적 해석을 가능하게 한다.

왼쪽 열의 그래프들은 각 인스턴스가 Class 0(정상 게시글)과 Class 1(마약 거래 게시글)로 분류될 확률을 보여준다. Instance 0은 높은 확률(약 0.99)로 정상 게시글로 분류되었으며, Instance 1과 2는 모두 높은 확률(약 0.99)로 마약 거래 게시글로 분류되었다.

오른쪽 열의 그래프들은 각 인스턴스의 분류에 영향을 미친 상위 10개 특성(단어)들의 중요도를 나타낸다. 녹색 막대는 정상 게시글(Class 0)로의 분류에 기여하는 특성을, 빨간색 막대는 마약 거래 게시글(Class 1)로의 분류에 기여하는 특성을 나타낸다. 막대의 길이는 해당 특성이 예측에 미치는 영향력의 크기를 의미한다.

[그림 18]의 경우, ‘나만의’, ‘기흥구’, ‘맛스타워요’ 등의 단어가 정상 게시글로의 분류에 강한 영향을 미쳤다. 반면, ‘미시맘’, ‘커피에’, ‘토론토가’ 등

은 마약 거래 게시글의 특성을 나타내는 것으로 분석되었다.

[그림 19]에서는 ‘붓’, ‘소개’, ‘토에 가능합니다’ 등이 정상 게시글의 지표로, ‘티디파리고요’, ‘최고예요’, ‘실아요’ 등이 마약 거래의 지표로 작용했다. [그림 20]에서는 ‘안재용’, ‘시작하세요’, ‘조가신’ 등이 정상 게시글의 특성으로, ‘윌러드튼게요’, ‘편디파리고’, ‘빠른거래원함’ 등이 마약 거래 게시글의 특성으로 나타났다.

LIME과 SHAP 분석을 통해 모델이 높은 정확도를 달성한 핵심 요인들이 드러났다. 첫째, 모델은 “당일가능”, “에이드”, “20만원”과 같은 일상 거래 용어들이 특정 맥락에서 마약 거래를 지시하는 은어로 활용되는 패턴을 성공적으로 포착했다.

둘째, “베스트셀러=0”, “세트판매=0”처럼 정상 거래 용어의 부재도 마약 거래 식별에 중요한 지표로 작용했다. SHAP 분석은 추가로 “토에 가능합니다”, “편디파리고” 등 특이한 단어 조합이 마약 거래 식별에 결정적 역할을 한다는 점을 보여줬다.

특히 SHAP의 경우 정상 게시글(0.99)과 마약 거래 게시글(0.99) 모두에서 높은 신뢰도의 예측을 보여주어, 모델이 이러한 언어적 패턴들을 안정적으로 학습했음을 입증했다. 두 모델의 분석 결과를 종합하면, 텍스트의 맥락과 단어 조합을 복합적으로 고려하는 접근이 마약 거래 탐지의 정확도를 크게 향상시켰음을 알 수 있다.

11. 소결

SNS상의 마약 거래 관련 텍스트는 일상적 단어를 은어로 활용하는 특징을 보인다. 분석을 위해 마약 거래 의심 게시글 1,560개와 정상 게시글 1,560개를 수집했다. 수집된 데이터는 한국어, 영어, 특수기호, 이모티콘을 포함하며 시간적 패턴과 사용자 행동 패턴 등 다양한 메타데이터를 포함한다. 특히 ‘아이스’, ‘떨’, ‘러쉬’, ‘크리스탈’ 등의 은어가 대유,

비유, 쌍관, 묘사의 네 가지 언어적 패턴으로 사용 되는 것이 특징이다.

워드 클라우드와 TF-IDF 분석 결과, 마약 거래 의심 게시물에서는 ‘직거래’, ‘배송’, ‘스타’, ‘아이스’ 등의 단어가 두드러지게 나타났다. 정상 게시물의 TF-IDF 점수는 20~60 사이인 반면, 마약 거래 의심 게시물의 은어들은 70~100 이상의 높은 점수를 기록했다. 토픽 모델링을 통해 정상 게시물과 마약 거래 의심 게시물 간의 주제적 특성 차이가 명확히 구분되었다. Word2Vec 분석에서는 마약 거래 의심 게시물의 단어들에 더 조밀한 의미 군집을 형성하고, 해시태그를 통한 은밀한 코드화가 이루어지는 것이 확인되었다.

로지스틱 회귀모델(99.89%), K-최근접 이웃 모델(99.68%), 의사결정나무 모델(98.61%)을 활용한 이상탐지 실험에서 매우 높은 정확도를 달성했다. LIME 분석을 통해 마약거래 게시글이 ‘당일가능’, ‘베스트셀러’, ‘특가전’ 등 일반 상거래의 특징을 의도적으로 모방하는 패턴이 발견되었다.

SHAP 분석에서는 일상적 단어들에 특정 맥락과 조합을 통해 마약 거래를 암시하는 은어로 활용되는 것이 확인되었다.

분석 결과를 종합하면, SNS상의 마약 거래는 일상적 단어를 은밀하게 활용하는 특수한 언어 패턴을 형성하고 있다. 이러한 패턴은 자연어 처리와 머신러닝 기법을 통해 98% 이상의 정확도로 탐지가 가능하다. 하지만 마약 거래 텍스트가 일반 상거래의 특징을 교묘하게 모방하고 있어, 단순한 키워드 매칭이 아닌 문맥을 고려한 복합적 분석이 필요하다. 특히 해시태그, 이모티콘, 특수문자 등 SNS의 고유한 특성을 활용한 은밀한 소통 방식에 대한 지속적인 모니터링이 요구된다.

IV. 결론

1. 결론 및 시사점

본 연구는 SNS 상에서 급증하는 마약 거래와 진화하는 은어 사용에 대응하기 위한 기술적 해결방안을 제시하는 것을 목적으로 한다. 특히 일상적 단어를 은어로 활용하는 SNS 마약 거래의 특수성을 고려하여, 자연어 처리와 머신러닝 기법을 결합한 효과적인 탐지 방안을 도출했다.

연구의 주요 결과는 다음과 같다. 첫째, 마약 거래 의심 게시물 1,560개와 정상 게시물 1,560개를 분석한 결과, ‘아이스’, ‘떨’, ‘러쉬’, ‘크리스탈’ 등의 은어가 대유, 비유, 쌍관, 묘사의 네 가지 언어적 패턴으로 체계화되어 있음을 확인했다. 이는 마약 거래 은어가 단순한 단어 변형이 아닌 정교한 언어 체계를 구축하고 있음을 시사한다.

둘째, 워드 클라우드와 TF-IDF 분석을 통해 마약 거래 텍스트의 특징적 패턴을 정량화했다. 마약 거래 의심 게시물에서는 ‘직거래’, ‘배송’ 등의 단어가 70~100의 높은 TF-IDF 점수를 보인 반면, 정상 게시물은 20~60 사이의 상대적으로 낮은 점수를 기록했다. 특히 Word2Vec 분석 결과는 마약 거래 관련 단어들에 조밀한 의미 군집을 형성하며, 해시태그를 통한 은밀한 코드화가 이루어지고 있음을 보여준다.

셋째, 머신러닝 기반의 이상탐지 모델들은 98% 이상의 높은 정확도를 달성했다. 특히 로지스틱 회귀모델은 99.89%의 정확도를 보여, 텍스트 기반 마약 거래 탐지의 높은 기술적 실현 가능성을 입증했다. LIME와 SHAP 분석을 통해 마약 거래 텍스트가 일반 상거래의 특징을 교묘하게 모방하는 패턴도 확인되었다.

본 연구의 기대효과는 다음과 같다. 첫째, 마약 거래 탐지의 정확도와 효율성이 크게 향상될 것이다. 특히 실시간 모니터링이 가능해져 거래가 이루어

어지기 전에 예방적 조치를 취할 수 있다. 둘째, 급증하는 청소년 마약 사범 문제에 선제적 대응이 가능하다. 2023년 10대 마약사범이 전년 대비 207% 증가한 현 상황에서, 이는 매우 중요한 의미를 갖는다. 셋째, 수사기관의 인력과 자원을 보다 효율적으로 활용할 수 있게 된다.

더욱이 본 연구에서 개발된 방법론은 성매매, 불법 도박, 보이스피싱 등 은어를 활용한 다양한 사이버 범죄 탐지에도 확장 적용이 가능하다. 이는 우리 사회의 사이버 범죄 대응 역량을 전반적으로 향상시키는 데 기여할 것으로 기대된다.

2. 한계점 및 보완 방향

본 연구의 한계와 향후 보완 방향은 다음과 같다. 첫째, 현재의 분석은 정적인 데이터셋을 기반으로 하고 있어, 빠르게 진화하는 마약 은어의 동적 특성을 완전히 반영하지 못한다. 이를 보완하기 위해 실시간 데이터 수집 및 분석 시스템의 구축이 필요하다.

둘째, 최근 등장한 생성형 AI 기술은 마약 거래 탐지 시스템에 새로운 도전과제를 제시한다. GPT와 같은 대규모 언어 모델은 기존의 마약 거래 패턴을 학습하여 더욱 정교한 위장 텍스트를 생성할 수 있다. 이에 대응하기 위해서는 생성형 AI 기반의 마약 거래 텍스트를 식별할 수 있는 고도화된 탐지 알고리즘의 개발이 필요하다.

셋째, 다크웹이나 메신저 등 다양한 플랫폼으로 확장된 마약 거래에 대한 통합적 분석이 부족하다. 플랫폼별 특성을 고려한 차별화된 탐지 전략과 함께, 플랫폼 간 연계 분석을 통한 종합적 탐지 시스템의 구축이 요구된다.

넷째, 현재의 모델은 텍스트 데이터에만 초점을 맞추고 있으나, 실제 마약 거래는 이미지, 이모티콘, 동영상 등 다양한 멀티미디어 요소를 활용한다. 멀티모달 분석 기법을 도입하여 텍스트 외 다양한 데

이터 유형에 대한 탐지 능력을 강화할 필요가 있다.

다섯째, 딥러닝 기반의 자기지도학습(Self-supervised Learning) 모델을 도입하여 레이블이 없는 대량의 SNS 데이터에서도 효과적으로 마약 거래 패턴을 학습할 수 있는 방법론의 개발이 필요하다. 특히 사전학습된 언어 모델을 마약 거래 도메인에 특화시키는 전이학습 기법의 적용을 고려해야 한다.

여섯째, 탐지 시스템의 설명가능성(Explainability)을 더욱 강화할 필요가 있다. 현재의 LIME과 SHAP 분석은 개별 예측에 대한 설명을 제공하지만, 모델의 전반적인 의사결정 과정에 대한 더 깊은 이해와 해석이 요구된다. 이는 수사기관의 실무 활용성을 높이고, 법적 증거로서의 가치를 확보하는 데 핵심적인 요소가 될 것이다.

일곱째, SNS 모니터링 과정에서 발생할 수 있는 프라이버시 침해와 과도한 검열의 우려가 있다. 불법 거래 탐지라는 공익적 목적과 개인의 기본권 보호 사이의 균형이 필요하며, 이를 위한 법적·제도적 장치 마련이 요구된다. 특히 일상적 단어를 활용한 거래 은어의 특성상, 정상적인 SNS 활동이 오탐지될 위험을 최소화하는 방안이 필요하다.

마지막으로, 마약 거래 탐지 시스템의 운영 과정에서 수집되는 개인정보의 보호와 관리에 대한 구체적인 지침이 필요하다. 이를 위해 데이터 수집 범위의 명확한 설정, 수집된 정보의 안전한 보관과 폐기 절차, 그리고 접근 권한 관리 등 체계적인 개인정보 보호 체계 구축에 대한 논의가 필요하다.

참고문헌

- 김병진, 이상준 (2019). 웹 사이트 간 연결 관계를 이용한 유해사이트 판별방법 개선 및 화이트 리스트 구축. 정보과학회컴퓨팅, 25(10), 506-510.
- 김세민, 전재현, 민현석, 노용만, 한승완, 최병철 (2010). 유해 이미지 분류 성능 개선을 위한 이중 피부 화소 검출을 이용한 인체 검출. 한국방

- 송미디어공학회 학술발표대회 논문집, 82-85.
- 김종훈, 박영섭 외 2명 (2005). 은어·비속어·직업어. 집문당.
- 김홍석 (2011). 은어(隱語)와 우리말의 세계, 글누림
- 민순호 (2017). 교육 서비스 플랫폼에서 유해 멀티미디어 비디오와 이미지 차단 기술 연구. 한국정보기술학회논문지, 15(7), 131-138.
- 송철환, 유성준 (2006). Support Vector Machine을 이용한 유해 이미지 분류. 한국정보과학회 학술발표논문집, 33(2C), 218-221.
- 신정훈 (2013). 웹 사이트간 연결 관계를 이용한 유해사이트 판별 방법. 박사학위논문, 숭실대학교 대학원.
- 안상선 (2024). 금융 분야의 이상탐지 모델 보안을 위한 통계 기반의 적대적 데이터 대응 방안- 실증적 분석과 시뮬레이션 결과. 미래사회, 15(3), 147-166.
- 안정근, 최은정, 조국인 (2013). 은어사용의 사회언어학적 고찰-세대별, 성별 차이를 중심으로. 건지인문학, 9, 285.
- 윤현식, 윤영호, 박현재(2020), 머신러닝을 통한 SNS 상의 성매매 알선 홍보 글탐지의 효율성 제고방안 연구. 한국지역정보화학회지, 23(3), 43-65
- 장정현, 나스리디노프 아지즈 (2017). 유해 해시태그 비율 기반의 유해 정보 판단 및 수집 시스템. 한국정보과학회 학술발표논문집, 273-275.
- 정규철, 이진관, 이태현, 박기홍 (2004). 문자 기반 유해사이트 판별 기법. 컴퓨터교육학회 논문지, 7(5), 83-91.
- 정병호, 임동훈 (2016). RHIFE 플랫폼에서 빅데이터 로지스틱 회귀를 위한 학습 알고리즘. 한국데이터정보과학회지, 27(4), 911-923.
- 조동욱, 김지영 (2004). 음란 유해사이트 차단을 위한 음향 신호 처리 및 분석. 한국콘텐츠학회논문지, 4(2), 1-6.
- 최태석 (2022). 한·중 마약 은어 비교 분석. 석사학위논문, 한국외국어대학교 대학원.
- 한병우, 윤지원 (2016). 단어 조합 검색을 이용한 불법·유해정보 탐지 기법. 정보보호학회논문지, 26(2), 397-404.
- Christopher J.Rhodes, *The Use of Open Source Intelligence in the Construction of Covert Social Networks*, Counterterrorism and Open Source Intelligence, 2011
- Lundberg, S., & Lee, S. I. (2017). *A Unified Approach to Interpreting Model Predictions*. arXiv:1705.07874v2.
- Neri, F., pettoni, M., (2008). *Stalker, A Multilingual Text Mining Search Engine for Open Source Intelligence*, Information Visualisation., IV '08. 12th International Conference 314-320.
- 뉴스시 (2020.01.15). 서울시 인터넷 시민감시단, 성매매 광고 7만 6,600건 감시.
- 이데일리 (2022.10.09). 아이스작대기 검색하니 마약판매글 주르륵... 삭제엔 35일 걸려.
- 전자신문 (2022.05.02.). 마약매매까지... 코인정보방 불법 온상 방치.
- 중앙일보 (2020.01.16). SNS “애인대행” 광고 정체는 성매매... 시민 신고로 딱걸렸다.
- 한겨레신문 (2019.06.20). 인공지능·딥러닝 이용한 디지털 성범죄 근절 추진.

투고일자: 2024. 12. 31.

심사일자: 2025. 1. 24.

게재확정일자: 2025. 2. 12.

Algorithm-Based Analysis of Drug-Related Transactions in Social Media Text: Focusing on Pattern and Context Analysis

Ahn, Sang-Sun

Kookmin University

This study proposed effective detection methods for responding to increasing drug transactions on SNS platforms and for continuously evolving slang usage. Using mixed data from suspected drug transaction and normal posts, a systematic analysis was performed using natural language processing and machine learning techniques. The analysis revealed four linguistic patterns in SNS drug transaction texts: metaphors, similes, double entendres, and descriptions. Through TF-IDF and Word2Vec analyses, we confirmed that these texts formed dense semantic clusters and utilized covert coding through hashtags. In the anomaly detection model experiments, the logistic regression model achieved 99.89% accuracy, and LIME and SHAP analyses revealed patterns in which drug transaction texts subtly mimicked the characteristics of general commerce. This study is significant for systematically identifying the linguistic characteristics of SNS drug transactions proposing effective detection methods, providing practical countermeasures for online drug trafficking issues. It is expected to be applicable for detecting various types of cybercrime

Keywords: Artificial Intelligence, Drug Trafficking, Natural Language Analysis, Cybercrime, Machine Learning