

SECaaS 기반 Multi-Modal Spam(Scam) Filtering Platform

클라우드 웹 기반에서 멀티모달 데이터를 정교하게 분석하여 스팸(스캠) 여부를 판별하고 결과를 제공하는 지능형 플랫폼

적용 분야



이메일 보안



클라우드 웹 서비스



AI 보안



멀티 모달 데이터 분석

연구 목적

- 스팸(스캠) 메일의 내용이 정교해지며 다른 형식의 데이터 혼합 형태로의 스팸(스캠) 메일 전파가 증가함에 따라, 정확하고 효율적인 스팸(스캠) 판별 필요성 대두
- 단순한 판별 결과 제공을 넘어, 판단 근거에 대한 분석을 통한 결과의 투명성과 신뢰성 확보 필요성 대두

As-Is 단일 형식의 데이터 위주 탐지로 인해 다중 데이터 형식 메일에 대한 효과적인 탐지 부족 및 탐지 결과에 대한 근거 설명 부족

To-Be 정형 데이터와 비정형 데이터를 함께 스팸(스캠) 필터에 활용함으로써 다중 데이터를 활용한 분석 및 XAI를 활용하여 탐지 결과에 대한 설명 가능한 분석 결과 제공

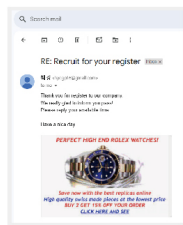
연구 내용

클라우드 웹 기반 멀티모달 스팸(스캠) 필터링 수행

스팸(스캠) 필터 | 기계학습 기반 스팸 필터 (예: Naive Bayes와 Support Vector Machine)를 다중으로 활용하여 각 필터에 따른 다양한 분석 결과를 제공

멀티 모달(Multi-Modal) 데이터 스팸(스캠) 판단 수행 | 이메일의 텍스트와 이메일 내에 첨부된 데이터를 추출하여 서로 다른 유형(텍스트, 이미지, 음성 등)의 데이터를 스팸(스캠) 분석에 함께 활용

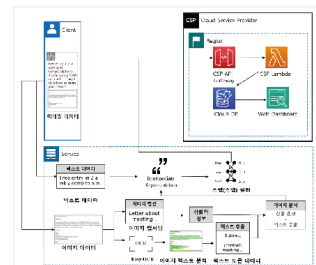
판단 결과 보고 | XAI(eXplainable AI)를 활용하여 이메일 데이터의 스팸(스캠) 판단 여부를 단어별 의심 확률 그래프와 하이라이트 형식으로 제공



이메일 불러오기



1. 플랫폼 동작 결과 분석 화면



2. 플랫폼 아키텍처 설계도

기술 경쟁력

기존기술

- 정형 데이터(텍스트) 분석에 의존하여 비정형 데이터(이미지, 음성 등)의 효과적인 스팸(스캠) 탐지 한계
- 스팸(스캠) 분석 결과에 대한 상세 근거 분석이 제공되지 않는 경우가 대다수를 차지함

기술적 한계

- 단일 데이터 : 텍스트 형식 위주의 분석 수행
- eXplainable AI 부재 : 스팸(스캠)으로 판별된 결과와 확률만을 결과로 알려주며 판단 근거를 제공하지 않음

기술 차별성

- 정형 데이터(텍스트)와 비정형 데이터(이미지, 음성 등)를 함께 분석에 활용하여 스팸(스캠) 탐지
- XAI(eXplainable AI)를 활용하여 스팸(스캠) 분석 결과에 대한 설명 가능한 결과 제공

기술적 우위

- 멀티 모달 데이터 : 텍스트 뿐 아니라 첨부된 다른 형식의 데이터를 함께 고려하여 분석
- eXplainable AI : 결과에 대한 과정 분석을 통해 스팸(스캠) 판단의 근거를 함께 제공

지식 재산권

발명의 명칭

Zone 유사도 기반 멀티 모달 스팸 필터링 방법 및 장치

출원(등록)번호

10-2024-0158354

출원(등록)일자

2024. 11. 08.