


프로젝트 소개

1. 팀 소개

팀명		노네임드	
팀원		이름	소속
멘토		정좌연	LG CNS
멘티	팀장	조상혁	엔코어 Playdata
	팀원1	문주혁	엔코어 Playdata
	팀원2	조동수	엔코어 Playdata
팀원 사진			

2. 프로젝트 정보

프로젝트 명	k-사이버 시큐리티 챌린지 비정상 이메일 분석
기술분야	빅데이터 분석 ■ python ■ 머신러닝
주제영역	■ 보안
수행기간	예선 : 20.10.01 ~ 20.10.30 / 본선 : 20.11.10 ~ 20.11.25
프로젝트 소개	비정상 이메일(스팸메일, 악성메일 등) 빅데이터 분석을 통한 공격패턴 및 특징을 도출함.

도전 과제	1. 비정상 메일 데이터 간의 차이를 분석하여 비정상 메일의 속성 및 특징을 분류 2. 송신 그룹의 활동에 대해 이상 징후를 탐지하여 프로파일링 3. 분석결과를 시각화								
프로젝트 진행	1. eml 파일을 분석을 위해 csv로 변환함 2. light gbm, 비지도 clustering 등 머신러닝 분석툴을 이용하여 악성메일 feature 추출 3. 추출한 주요 feature별로 추출한 분석 결과를 시각화하여 표현함 ex) 메일 사이즈, 그림 수, 언어, 메일 안 텍스트 분석 4. 진행 내용을 정리하여 보고서 작성								
적용기술	<table border="1"> <tr> <td>데이터 수집</td><td>비정상 이메일을 eml 형식으로 제공</td></tr> <tr> <td>데이터 정제</td><td>eml을 파일을 인코딩 및 csv 파일에 주요 정보 적재 # 활용 : Python</td></tr> <tr> <td>데이터 분석</td><td>1. 비지도 학습 clustering 방식을 이용하여, 비정상 이메일의 feature를 찾음 2. labeling된 데이터를 토대로, 머신러닝 지도학습을 통해 각 feature별 중요도 파악 # 활용 : Python</td></tr> <tr> <td>데이터 시각화</td><td>python의 pandas 라이브러리 및 R의 ggplot 라이브러리를 활용하여 분석결과 시각화</td></tr> </table>	데이터 수집	비정상 이메일을 eml 형식으로 제공	데이터 정제	eml을 파일을 인코딩 및 csv 파일에 주요 정보 적재 # 활용 : Python	데이터 분석	1. 비지도 학습 clustering 방식을 이용하여, 비정상 이메일의 feature를 찾음 2. labeling된 데이터를 토대로, 머신러닝 지도학습을 통해 각 feature별 중요도 파악 # 활용 : Python	데이터 시각화	python의 pandas 라이브러리 및 R의 ggplot 라이브러리를 활용하여 분석결과 시각화
데이터 수집	비정상 이메일을 eml 형식으로 제공								
데이터 정제	eml을 파일을 인코딩 및 csv 파일에 주요 정보 적재 # 활용 : Python								
데이터 분석	1. 비지도 학습 clustering 방식을 이용하여, 비정상 이메일의 feature를 찾음 2. labeling된 데이터를 토대로, 머신러닝 지도학습을 통해 각 feature별 중요도 파악 # 활용 : Python								
데이터 시각화	python의 pandas 라이브러리 및 R의 ggplot 라이브러리를 활용하여 분석결과 시각화								

3. 의사소통방법

- 주 1회 이상 주기적으로 오프라인 미팅을 통해 아이디어 공유 및 개발 현황 점검
- 오프라인 미팅이 불가할 시 줌을 통한 온라인 미팅으로 대체
- 카카오톡을 통한 온라인 피드백은 상시 진행