머신러닝 기반
악성 URL 판별 기법 개발

신을 찌르면? 푹신푹신

팀원

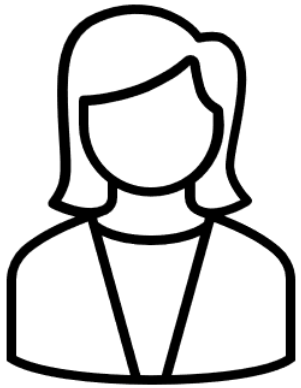
강수진
강민성
문동준
오현진
주현우

담당교수

양환석

목차

서론_팀원 소개



팀장
강수진
19학번

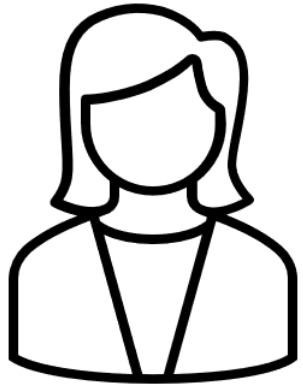
DGA
&
Flask

팀원
강민성
18학번

머신러닝
&
GUI

팀원
문동준
17학번

파이썬
&
백엔드



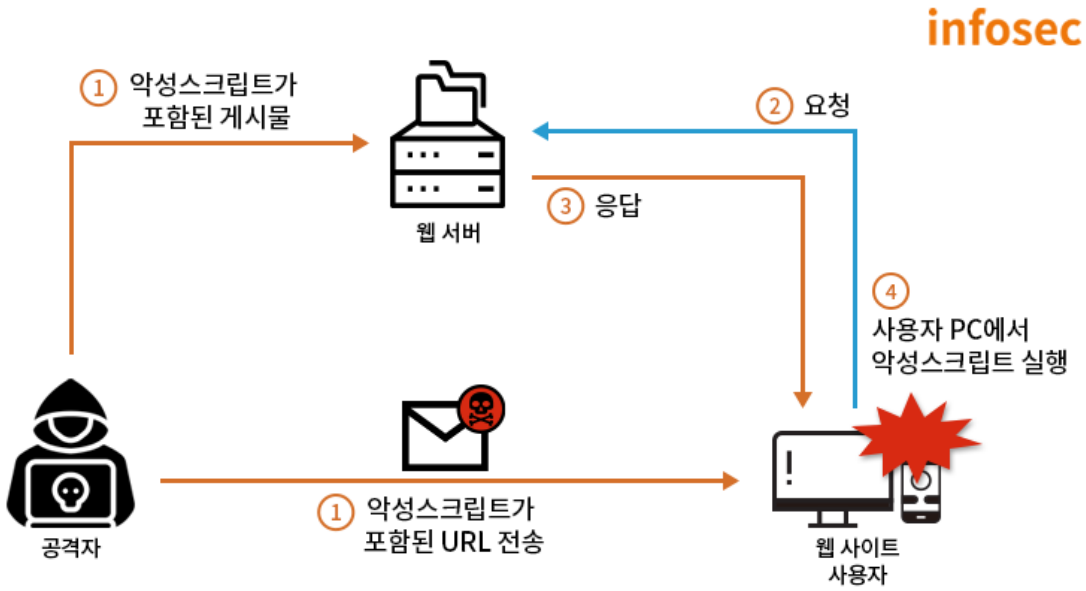
팀원
오현진
20학번

프론트엔드
&
백엔드

팀원
주현우
17학번

머신러닝
&
GUI

서론_프로젝트 배경



악성 URL은 계속 발전

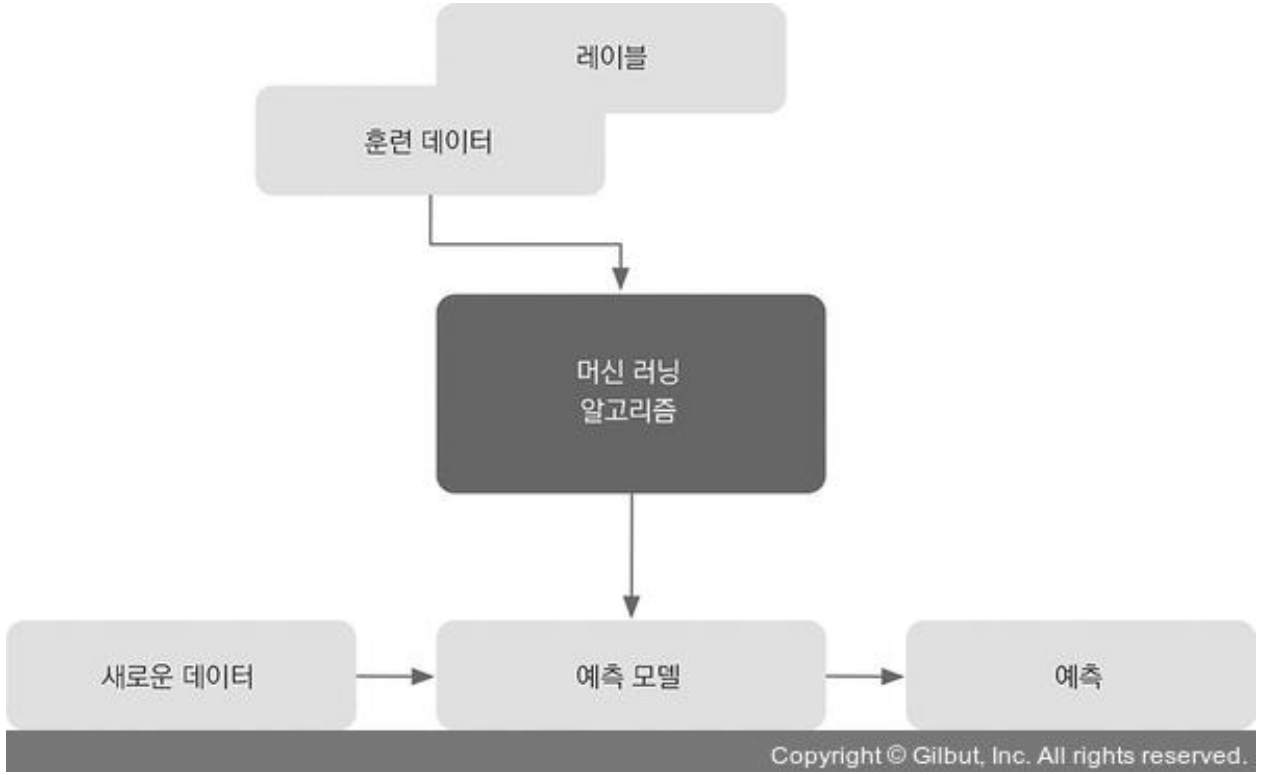
악성 URL로 인한 피해 방지 & 대응

유의미한 특징값 추출

Python 머신러닝 알고리즘과 DGA

Web과 GUI로 결과 확인

서론_계획안



파이썬 기반
Flask로 연결
API 사용

악성 URL feature 추출
->
추출 값을 알고리즘에 학습

본론_특징값 추출

CATEGORY	FEATURES	NUMBER
개수 기반	도메인 개수, http개수, https개수, .개수, //개수, -개수, @개수, www개수, =개수, _개수, ~개수, ?개수, &#%개수, 악성문자열 개수, 숫자 개수, 쿼리 개수, 쿼리 악성 문자열 개수	17개
길이 기반	url 길이, url path 길이, url netloc 길이, url tld 길이, 쿼리 길이	5개
존재 기반	쿼리 인코딩 유무, ip포함유무, 단축서비스 유무	3개
비율 기반	랜덤한 정도, 대문자 알파벳 비율	2개
도메인 기반	포트번호, 도메인 생성일~현재, 현재~도메인 만료일, 도메인 전체수명, 트래픽길이, abnormal유무	6개

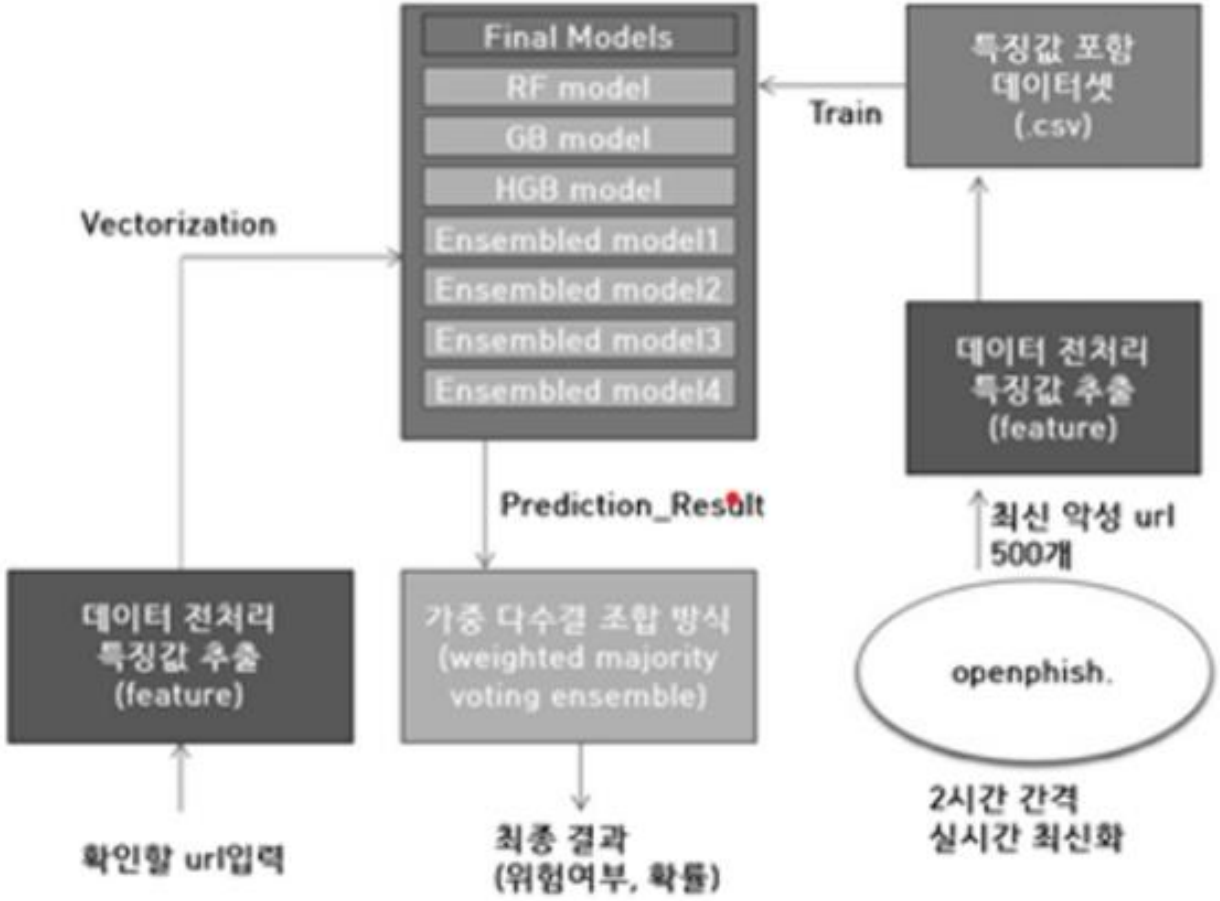
기존의 머신러닝을 활용한 악성 URL탐지 모델:

개수 기반, 길이 기반, 존재 기반, 비율 기반 등의 어휘적 특징

추가한 URL탐지 모델:

도메인 기반
도메인 수명, 트래픽 길이, 도메인 생성 일자 등

본론_머신러닝



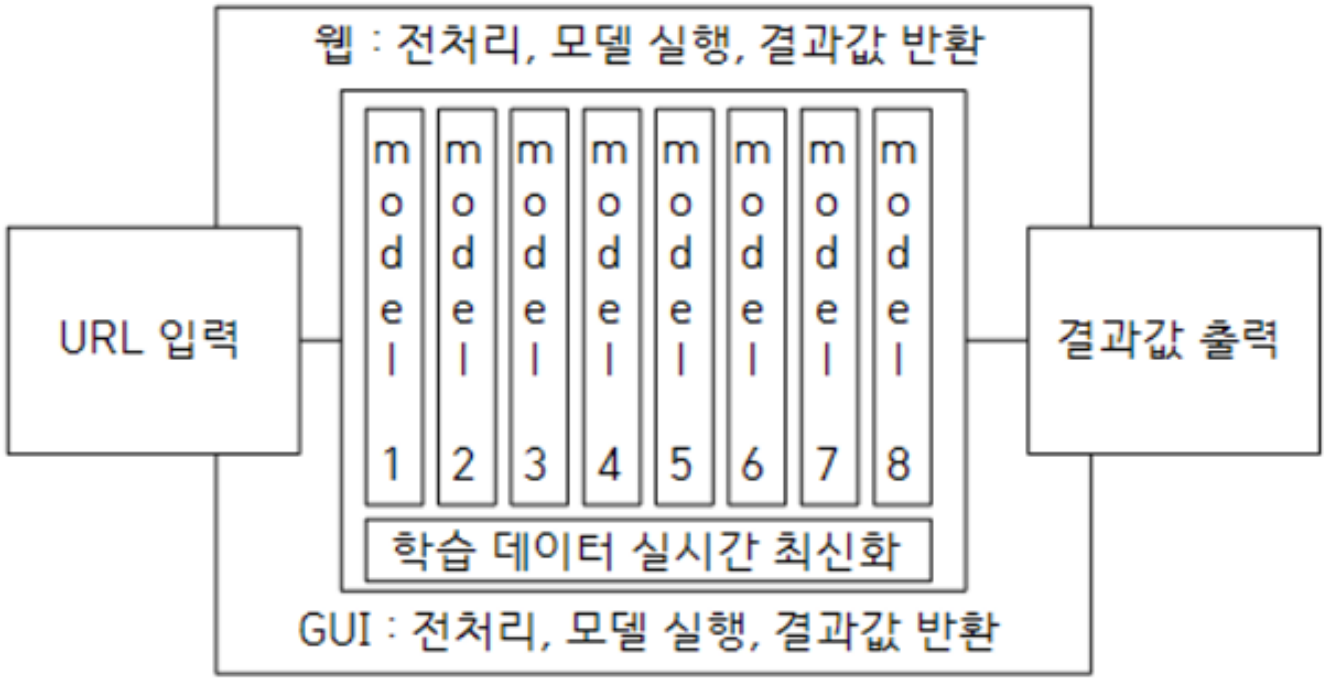
urllib 라이브러리: 도메인 정보 출력

csv 라이브러리: 출력된 특징값을
데이터화 하여 저장

pandas 라이브러리: 학습 가능한
형태로 저장

joblib 라이브러리: .h5 확장자 파일로 저장

본론_머신러닝



CNN을 활용해 총 8개의 모델을 기준으로 악성 여부 판단

특정 URL이 악의적인 목적하에 동적으로 생성되는 도메인인지 확인 과정 필요

최신 악성 URL을 전처리를 통해 모델에 학습

검사가 필요한 URL 입력

학습된 모델로 결과 확인

본론_DGA

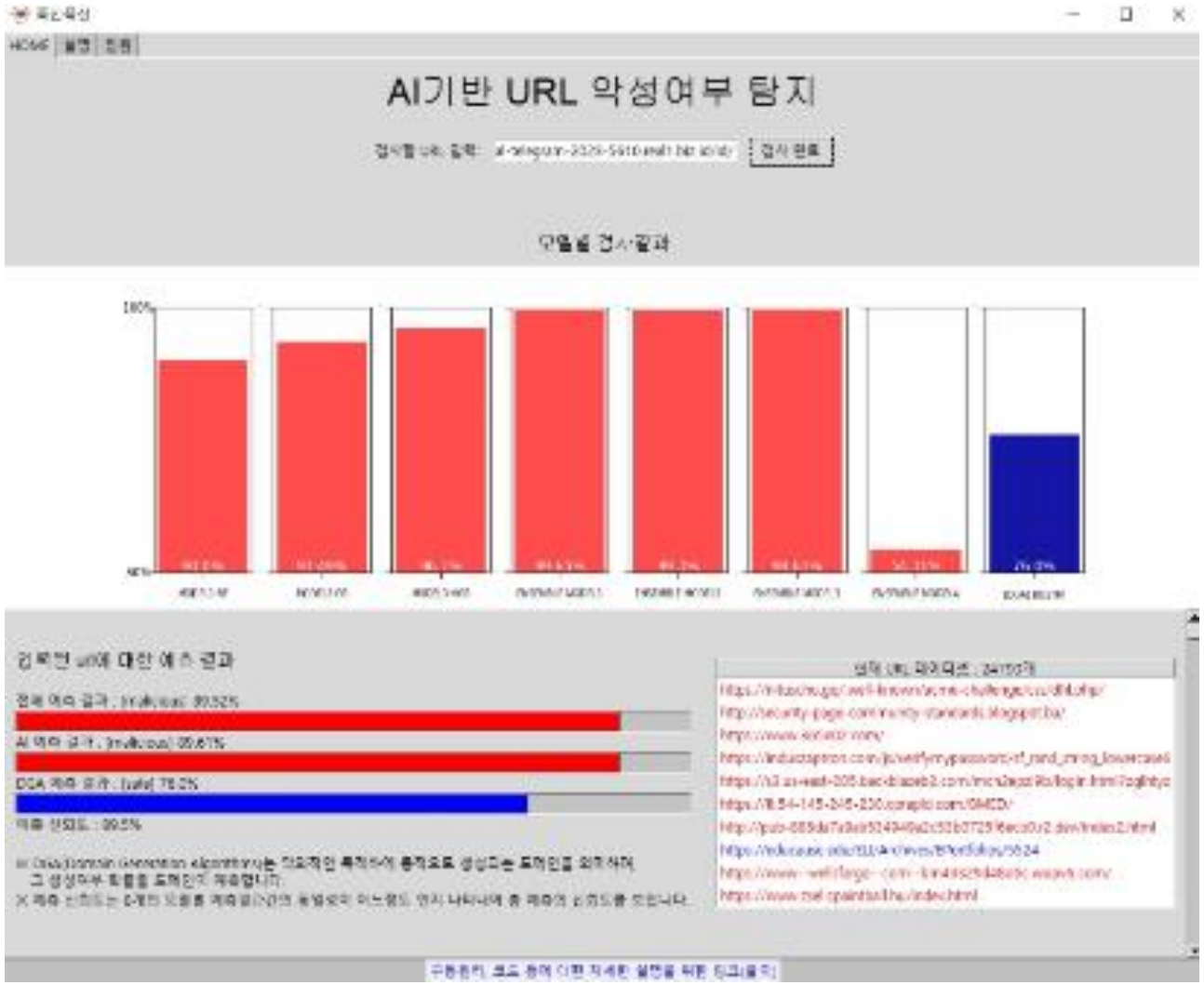
DGA	Description
CryptoLocker	DGA를 사용하여 C&C 서버에 연결하고, 암호화된 파일을 복원하기 위해 피해자에게 요구하는 도메인을 동적으로 생성
NewGoz	금융 부문을 타켓으로 하는 악성 코드로, DGA를 사용하여 악성 서버와 통신하고 금전적 이익을 추출하는 데 활용
GameOverDGA	DGA를 사용하여 봇넷과 통신하며, 금융 정보를 탕취하고 악성 활동을 숨기기 위해 동적 도메인을 생성
Nivdort	DGA를 활용하여 트로이 목마를 배포하고, 사용자의 개인 정보를 탕취하거나 다른 악성 코드를 설치
Necurs	대규모 스팸 및 악성 파일 배포를 위해 DGA를 사용하며, 다양한 악성 활용에 이용
Goz	금융 정보를 탈취하고 악성 활동을 숨기기 위해 DGA를 활용하는 악성 코드
Bamital	클릭 사기 및 광고 클릭 부정행위를 실행하며 DGA를 사용하여 도메인을 동적으로 생성하여 악성 활동을 수행

"남궁주홍, DGA 도메인 탐지를 위한 호율적인 딥러닝 모델"
국내석사학위논문 강원대학교 대학원, 2020, 강원도 논문

DGA :
Cryptolocker, newdoz, gameoverdga, Nivdort, necurs, goz, bamital

Non-DGA :
alexa, legit

결론_GUI



결과를 사용자의 가시성을 위해
GUI와 Web을 활용

URL의 입력
->
해당 URL을 모델에 입력
->
결과물 반환

검사 결과
AI 검사 결과
DGA 검사 결과
모델별 측정 결과
결과에 대한 신뢰도 항목

결론_Web

결론_성능

RF : 93.86%
GB : 90.74%
HGB : 90.39%
DT : 90.03%

스태킹 방식
RF-GB-ET-MLP-LR
DT-RF-KN-MLP-LR
RF-GB-MLP-AB-HGB

보팅 방식
RF-GB-HGB

Algorithm	Test Set	New Data Set	
	Accuracy	Benign Accuracy	Malicious Accuracy
DT	90.93%	75%	80%
KN	84.51%	77.5%	67.5%
GNB	50.62%	17.5%	95%
MLP	71.36%	52.5%	85%
RF	93.86%	90%	87.5%
LR	81.23%	55%	67.5%
AB	88.07%	90%	72.5%
GB	90.74%	87.5%	80%
ET	80.72%	92.5%	70%
HGB	90.39%	87.5%	77.5%
[S]RF-GB-ET-MLP-LR	94.18%	90%	85%
[S]DT-RF-KN-MLP-LR	93.95%	90%	87.5%
[S]RF-GB-MLP-AB-HGB	94.53%	90%	85%
[S]KN-GNB-MLP-RF-GB	93.41%	90%	87.5%
[V]DT-KN-GNB	96.88%		
[V]DT-MLP-RF	99.11%		
[V]KN-GNB-MLP	93.98%		
[V]DT-GNB-RF	99.11%		
[V]KN-MLP-GNB	92.08%		

결론_결과

최근 악성 **URL** 을 활용한 사이버 위협이 지속되고 있으며 새로운 패턴에 대해 예측하여 대응하기 위한 정보 보안 시스템의 중요성이 강조되고 있음

파이썬**(Python)** 환경에서 실시간 데이터 최신화, 유의미한 특징값들을 간소화된 과정으로 신속하게 출력하고 이를 다수의 머신러닝 알고리즘 혹은 다중 머신러닝 알고리즘에 학습시킨 후 고성능의 모델을 활용해 실제 웹 기반 악성 **URL**판별 서비스를 제공함

특징값 별 유효도 검사를 통해 기존의 특징 추출 과정을 보완 및 학습에 용이한 새로운 특징값을 확보하고 머신러닝 알고리즘뿐 아니라 딥러닝 알고리즘까지 접목시킴

감사합니다

Q&A