Guard QR

- 1831005 강동인
- 1971362 이준희
- 1971406 김진웅
- 2171061 강세빈

목차

프로젝트 선정 배경 – 1

- 최근 QR코드를 이용한 피싱의 일종인 ‘큐싱’이 국내에 확산되고 있음

QR코드 무심결에 찍었다가 탈탈 털린다…'큐싱' 사기 주의

메일 발신인 '안랩닷컴'으로 위장…'임금 수령 통지서'라며 가짜 QR코드로 개인 정보 빼내

장유미 기자

[자세히보기] SW개발 과정에도 품질인증 제도가 있다!! SW프로세스 품질인증(SP인증) 제도 소개

코로나19 팬데믹 이후 활용도가 높아진 'QR(큐알)코드'가 최근 피싱범죄 도구로 이용되고 있어 각별한 주의가 요구된다.

2일 안랩 시큐리티 인텔리전스 센터(ASEC)에 따르면 최근 종화인민공화국 재정부를 사칭한 큐싱 메일이 빠르게 확산되고 있는 것으로 드러났다. 큐싱이란 QR코드와 피싱(Phishing)을 합한 단어로, QR코드를 이용한 사기 수법을 말한다.

ASEC는 큐싱 메일이 '2024년 1분기 임금 수령 통지서'를 위장하고 있다고 밝혔다. 메일 본문에는 임금 보조금을 수령하려면 휴대폰을 이용해 QR 코드를 스캔하도록 유도하는 문구가 포함돼 있다. 큐싱 발신인은 발신자 메일 주소를 '안랩닷컴'으로 위장했다.

Qlik DAEKYO

차세대 분석플랫폼의 효율을 높이기 위한 Qlik 데이터 통합 솔루션

2024.6.27(목) 사전등록

MZ 세대 SNS '큐싱 주의보'

이용객 "QR코드 진한정도 달라"

큐싱 포함 피싱범죄 매년 증가세

QR로 스마트폰에 악성코드 심어

금융 정보 탈취·금전적 이득 취해

QR코드 찍었다가 1000만원 탈탈…큐싱사기 아시나요

윤진섭 기자 입력 2024.02.19.06:52 수정 2024.02.19.21:05

자영업자 A씨는 '소상공인 저금리 대출'과 관련한 메일을 받았습니다. 메일에는 대출 안내와 함께 금융사기 예방 앱 설치를 위해 QR코드를 촬영하라는 내용이 있었습니다.

스마트폰으로 QR코드를 촬영한 A씨. 이후 A씨의 휴대전화에는 악성 앱이 설치됐고, 개인정보가 유출돼 1000만원 넘는 금전적 피해를 당하게 됐습니다.

신용보증재단중앙회가 실제 접수된 사기 피해 사건으로 소개한 A씨의 사례는 전형적인 '큐싱(Qshing)' 피해 사례입니다

SQUARESPACE

Get your brand online with a Squarespace website.

Alternata Tailors

많이 본 뉴스

1 차 만들어도 안 팔려요..소비가 식었다

2 "여행비 받고 한 달 살기"..다음 달 13일까지 모집

3 파리 중학교에서 '한식 급식'

4 2980만원 내고 나도 집주인?..또 고개 드는 갭투자

5 "미술관 퇴거 소송 취하해라"..최 회장 꾸짖은 재판부

프로젝트 선정 배경 – 2

- 큐싱이란, QR코드로 악성 앱 설치를 유도해 개인정보와 금융정보를 빼내는 수법을 의미
- 스캔할 QR코드에 악성 앱 다운로드 URL을 몰래 숨겨두고, 사용자가 의심 없이 QR코드를 읽으면 자동으로 URL에 연결되어 악성 앱이 다운로드 되는 원리
- 악성 앱이 설치되면 피해자의 보안카드 번호 및 공인인증서 등 각종 개인정보가 유출되며, 스마트폰 설정을 바꿔 소액결제 인증문자를 우회하는 등 금전 탈취 문제까지 발생

프로젝트 선정 배경 – 3

- 이러한 큐싱으로 인한 피해를 방지하기 위해 기존의 QR 코드 스캐너에 딥러닝 기반 악성 URL 탐지 모델을 추가하여 QR 코드를 스캔하였을 때 스캔 된 URL이 악성인지 정상인지 판별하여 접속을 방지하는 QR Scanner 어플을 고안

프로그램 개요

QR 코드 스캔

추출한 URL

딥러닝 모델

정상

악성

바로 사이트 접속

사용자에게 경고 후
접속 여부 확인

접속 O

접속 X

앱 초기화면

부가기능

QR 코드 스캔

추출한 URL

악성인 경우

추출한 URL 사이트의

1) 입력 값 여부

2) 리다이렉션 여부

3) 데이터 전송 destination

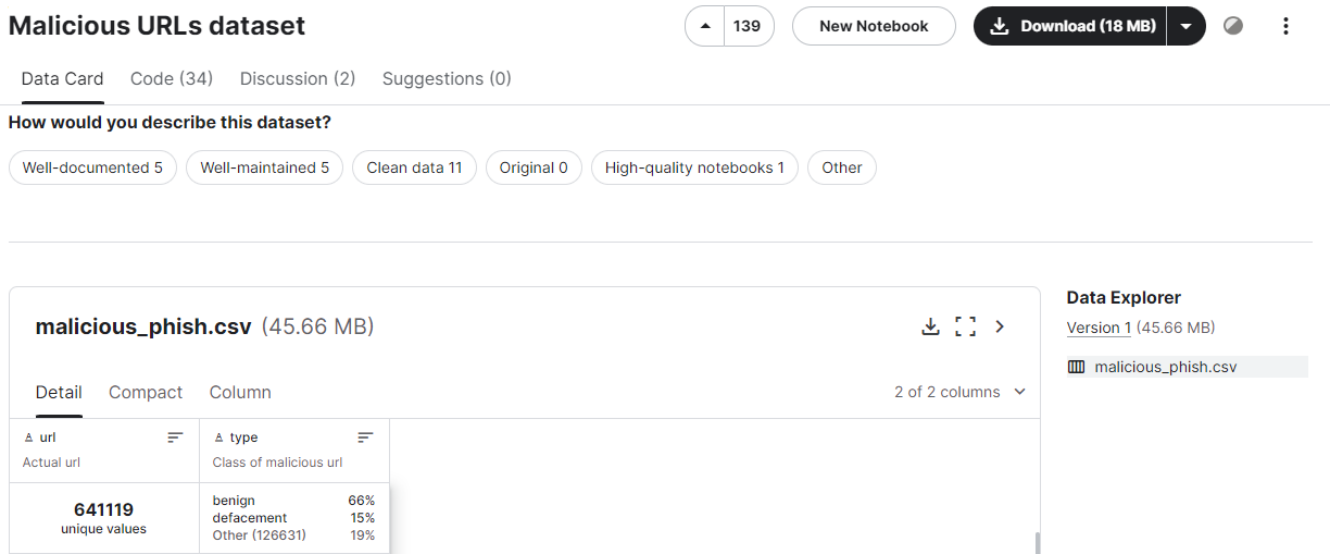
위 정보들을 사용자에게 제공

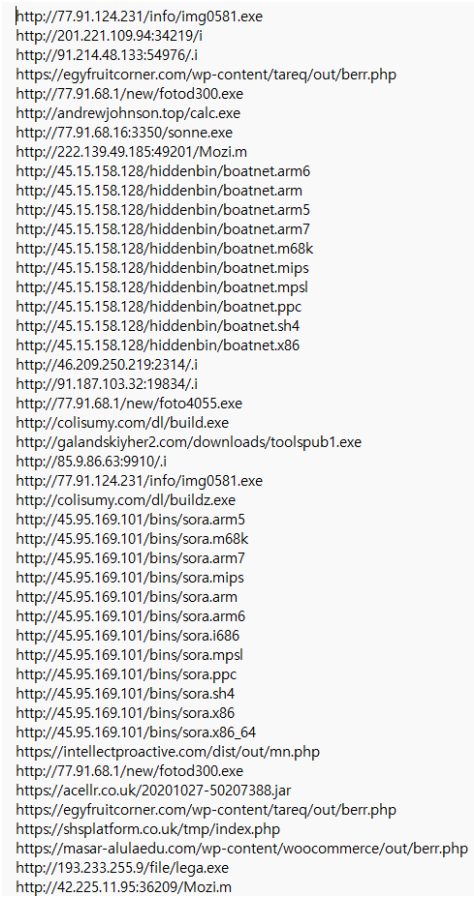
부가기능은 사용자가 on/off 할 수 있음

기술 스택

- **프로그래밍 언어**
  - Python, Java
- **프레임워크 및 라이브러리**
  - Tensorflow, Scikit-learn, Matplotlib, Whoisdomain
- **개발 도구 및 환경**
  - Visual Studio Code, Android Studio, Jupyter Notebook

핵심 기술 소개 – 사용한 데이터셋





기존 Kaggle 데이터셋
- 정상 : 423,139개
- 악성 : 217,980개

크롤링한 악성 URL
- 212,308개

두가지 데이터셋을 합쳐 악성 URL의 개수와 정상URL의 개수를 1:1 비율로 맞춤

이후 데이터 셋을 Train, Valid, Test 각각 7 : 1.5 : 1.5로 분할

핵심 기술 소개 – 추출한 특성

1) 접두어를 제외한 URL 주소의 길이
2) URL 단축 서비스 사용 여부
3) 정상적인 URL 형식 사용 여부
4) https 사용 여부
5) URL 주소내 ip 주소 포함 여부
6) URL 주소내 포함된 특수문자의 개수
7) Root domain의 길이
8) URL 주소내 이메일 주소 포함 여부
9) URL 주소내 파일 확장자 포함 여부
10) URL의 도메인 등록 기간
11) URL의 TTL

url	type	url_type	url_length	pri_domair	shorten_se	abnormal_	https	have_ip	root_doma	count_cha	root_doma	exist_email	exist_file_e	url_period	ttl
br-icloud.c	phishing	1	16	br-icloud.c	0	0	0	0	br-icloud	2	9	0	0	366	3600
mp3raid.cc	benign	0	35	mp3raid.cc	0	0	0	0	mp3raid	5	7	0	1	730	600
http://buzz	benign	0	111	buzzfil.net	0	1	0	0	buzzfil	8	7	0	1	416	300
espn.go.co	benign	0	45	espn.go.co	0	0	0	0	go	9	2	0	0	395	60
yourbittorr	benign	0	46	yourbittorr	0	0	0	0	yourbittorr	3	14	0	0	373	300
http://pash	defacemer	1	33	pashminac	0	1	0	0	pashminac	5	14	0	0	365	86400
allmusic.cc	benign	0	45	allmusic.cc	0	0	0	0	allmusic	3	8	0	0	3067	10800
corporatio	benign	0	62	corporatio	0	0	0	0	corporatio	5	15	0	0	365	300
myspace.c	benign	0	30	myspace.c	0	0	0	0	myspace	4	7	0	0	2229	60
quickfacts.	benign	0	44	quickfacts.	0	0	0	0	census	7	6	0	1	602	-1
nugget.ca/	benign	0	52	nugget.ca	0	0	0	0	nugget	5	6	0	0	325	43200
uk.linkedin	benign	0	46	uk.linkedin	0	0	0	0	linkedin	7	8	0	0	554	48
http://vnic.	defacemer	1	23	vnic.co	0	1	0	0	vnic	6	4	0	1	755	3600
baseball-re	benign	0	48	baseball-re	0	0	0	0	baseball-re	5	18	0	0	2696	300
192.com/a	benign	0	35	192.com	0	0	0	0	192	6	3	0	0	274	300
nytimes.co	benign	0	83	nytimes.co	0	0	0	0	nytimes	7	7	0	1	762	120
songfacts.c	benign	0	33	songfacts.c	0	0	0	0	songfacts	4	9	0	0	2810	3600
http://holly	benign	0	78	hollywoodl	0	1	0	0	hollywoodl	9	13	0	0	397	300

핵심 기술 소개 – 학습에 사용한 특성

FEATURE NAME	ODDS RATIO
url_length	0.97
shorten_service	1.08
abnormal_url	2.02
https	1.08
have_ip	1.05
count_characters	1.19
root_domain_length	1.05
exist_email	1.08
exist_file_extension	1.08
url_period	1.0
ttl	1.0

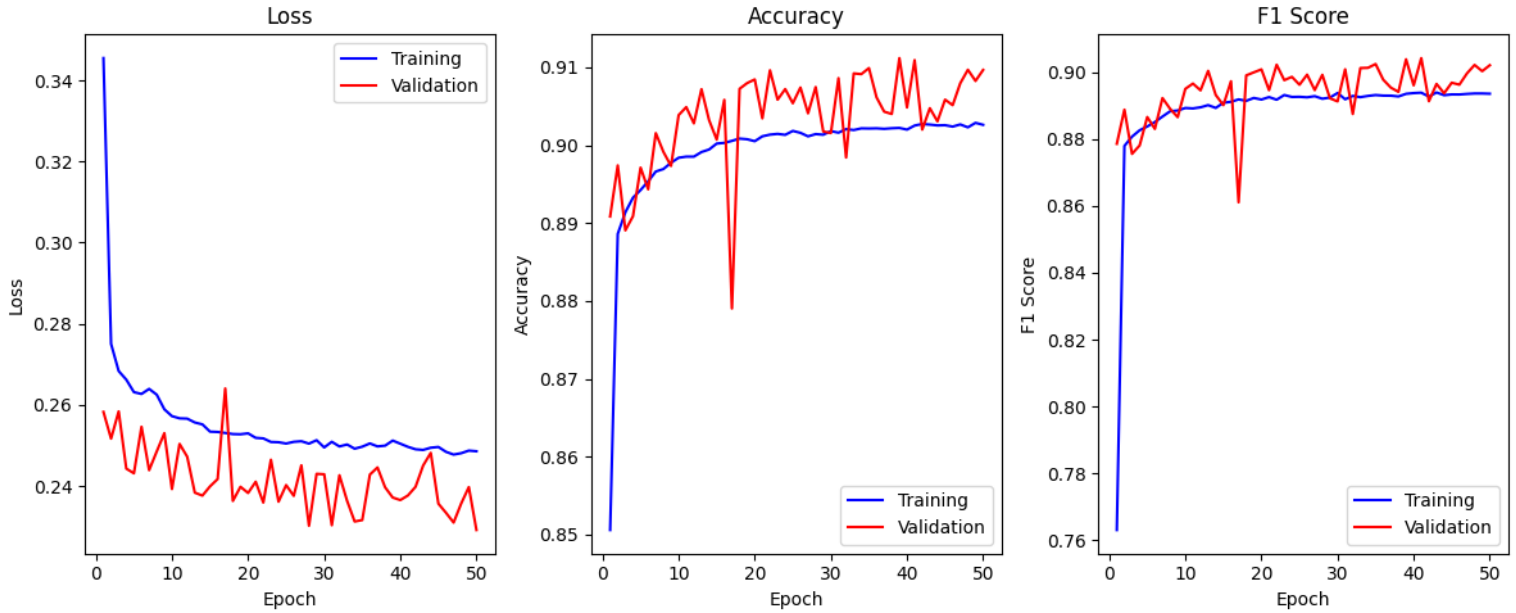
악성 URL 판별에 영향을 끼치지 않는 2개의 특성(url_period, ttl)을 제외한 나머지 9개의 특성을 모델 학습에 사용

핵심 기술 소개 – 사용한 모델

Model Architecture	Number of nodes
Dense Layer + Relu + Dropout(0.5)	128
Dense Layer + Relu + Dropout(0.5)	64
Dense Layer + Relu + Dropout(0.5)	32
Dense layer + Sigmoid	1

핵심 기술 소개 – 모델 학습 결과

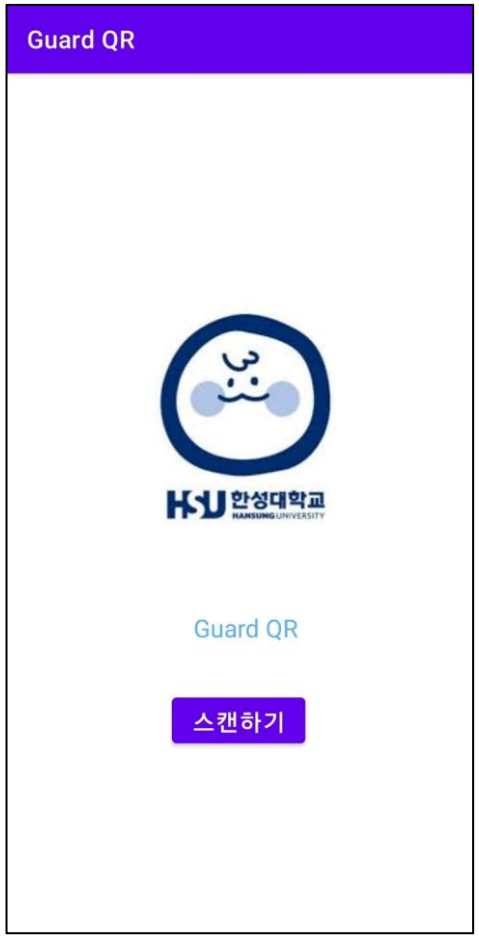
1) 학습 결과 그래프



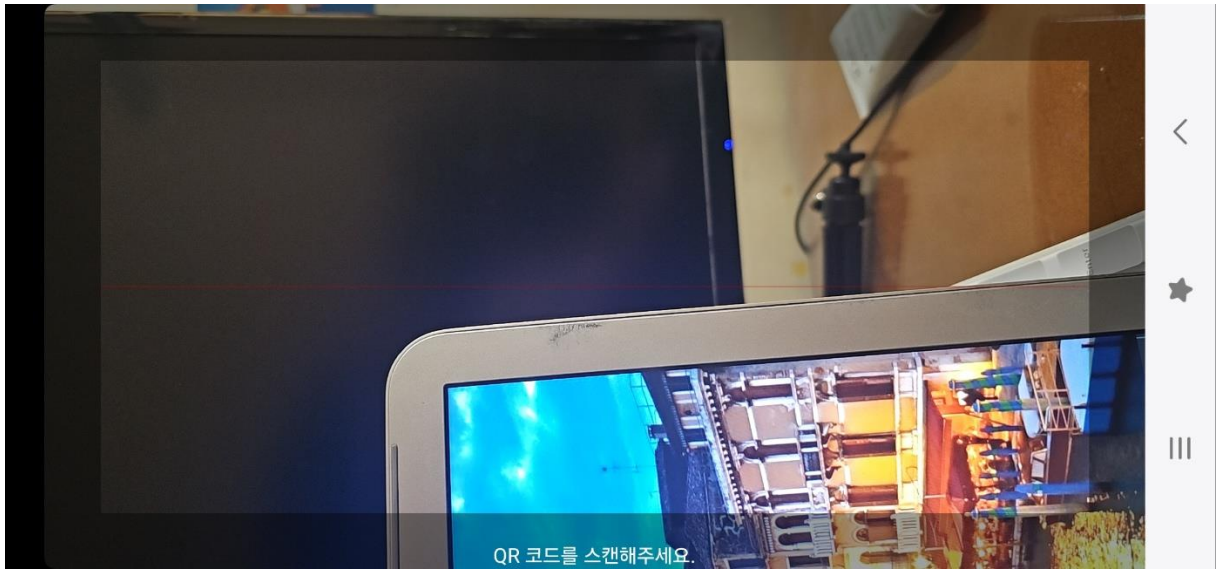
2) 테스트 데이터 기준 성능 측정 결과

Loss	Accuracy	Precision	Recall	F1 Score
0.2285	0.9097	0.9810	0.8313	0.9002

기능 시연 – 핵심 기능

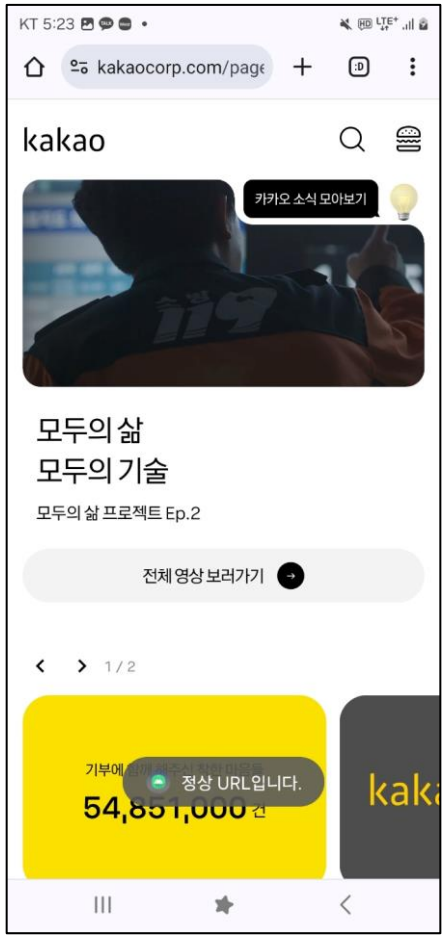


앱 초기화면

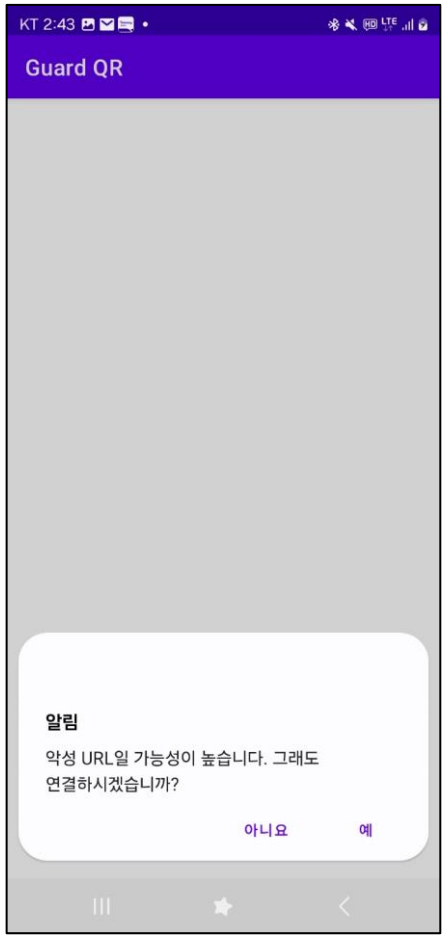


카메라가 켜진 화면

기능 시연 – 핵심 기능

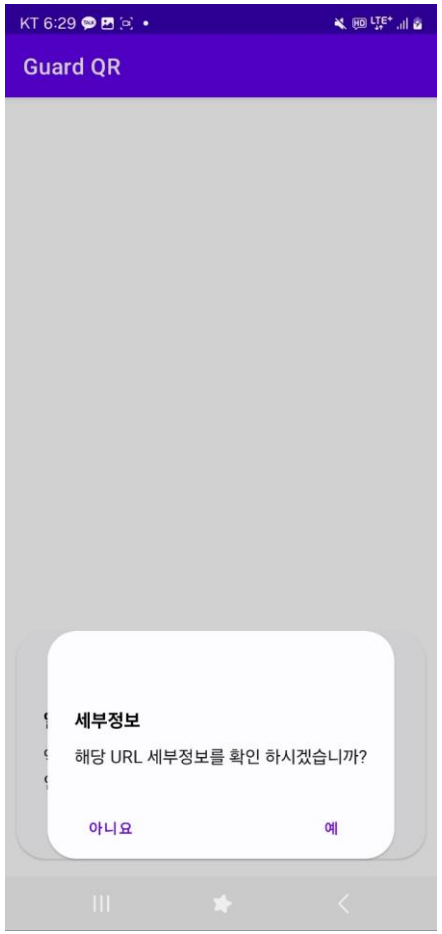


정상 URL인 경우

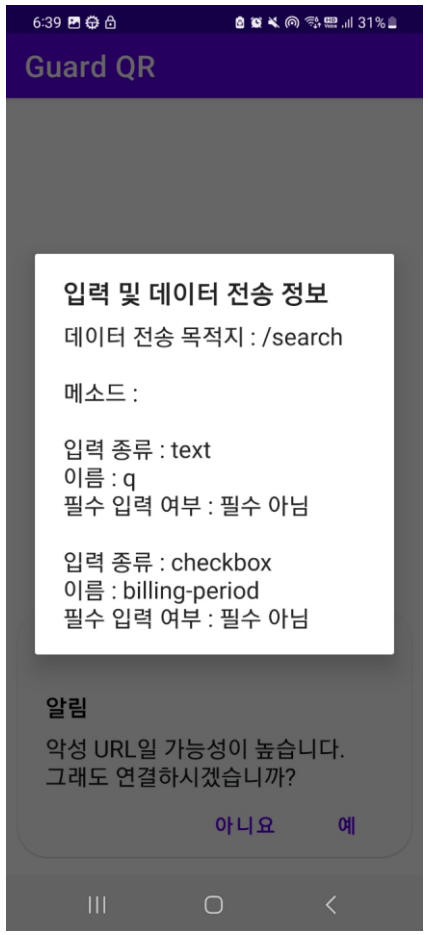


악성 URL인 경우

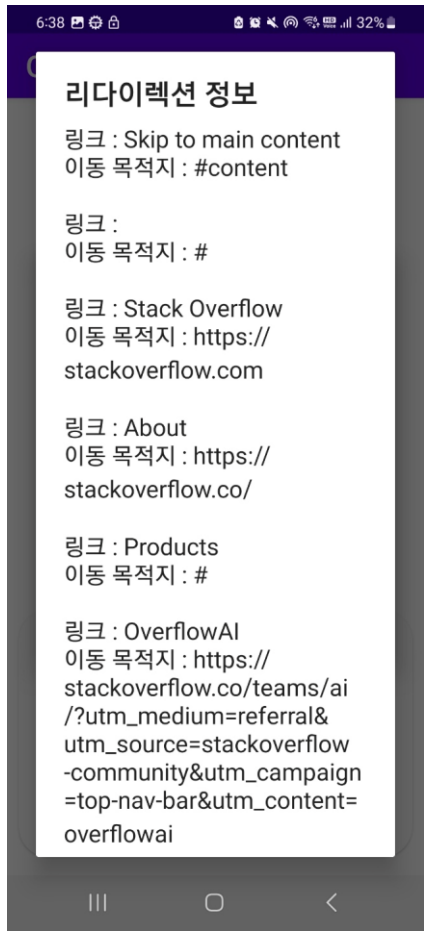
기능 시연 – 부가 기능



세부정보 선택



입력 및 데이터 전송 정보 출력



리다이렉션 정보 출력

한계점

- 최근에 구축된 URL 데이터셋이 부족하여 비교적 옛날에 구축된 URL 데이터셋을 사용하여 모델 학습을 진행

=> 최근에 유포되는 악성 URL을 이용하여 모델 학습을 진행하면 보다 더 좋은 성능을 기대할 수 있음

- 11개의 특성을 추출하였지만 악성과 정상을 구분 짓기에는 부족함.

=> 더 효과적인 특성을 추가하여 보완 가능

참고문헌

- 한채림, 윤수현, 한명진, 이일구. (2022). 머신러닝 기반 악성 URL 탐지 기법. 정보보호학회논문지, 32(3), 555-564.

- 권현, 박상준, 김용철. (2021). 뉴럴네트워크 기반에 악성 URL 탐지방법 설계. 융합정보논문지, 11(5), 30-37.

감사합니다