

딱걸렸조 졸업 프로젝트

AI로 활용한 URL

탐지 및 분석 웹 서비스

지도 교수: 유승재 교수님

팀장: 오준석

팀원: 권기범, 김응찬, 김현진

2025. 11. 04

목 차

프로젝트 개요

- 주제 선정 이유
- 팀 소개 및 역할 분담
- 프로젝트 추진 일정

프로젝트 진행

- 프로젝트 구성도
- 프로젝트 개발 환경
- 프로젝트 기능 구현

프로젝트 결과

- 웹 서비스 GUI
- 추후 계획
- 시연 영상

QnA

프로젝트 개요



프로젝트 개요: 주제 선정 이유

URL 링크를 통한 피싱 사례.. 여전히 많은 비중으로 사람들이 피해를 입고 있습니다.

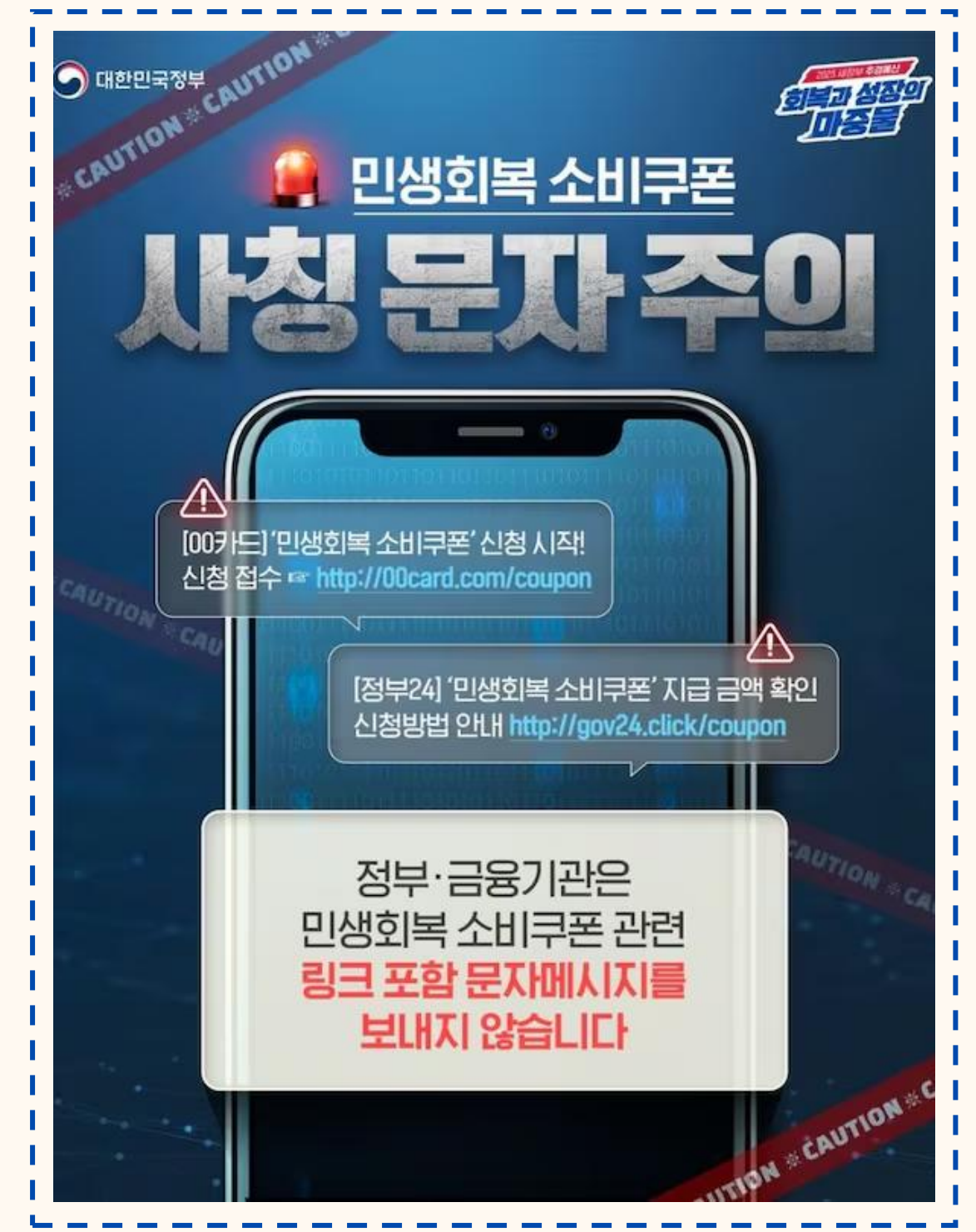
URL을 활용한 피싱 비율



악성 URL 탐지 수 증가

누리집에 따르면, 8월 한 달 동안 탐지된 악성 URL은 총 28만6천여 건에 달했다. 이는 7월에 탐지된 7만6천여 건보다 약 3.7배 증가한 수치다. 악성 URL 확산 규모가 폭발적으로 늘면서, 발령된 피싱주의보 건수도 동반 상승한 것으로 분석된다. 회사 측은 "피싱 공격자들이 국내 이용자를 대상으로 한 악성 URL 배포 빈도를 높이고 있으며, 공격 수법 또한 갈수록 지능화되고 있다"고 설명했다.

URL을 사용한 사칭 문자



팀원 구성과 각자 담당했던 역할에 대한 소개입니다.



팀장: 오준석

- 백엔드(총괄)
- 피쳐코드, ML 모델 개발
- 코드 통합



팀원: 김응찬

- 백엔드
- 피쳐코드, ML모델 개발
- 데이터 전처리



팀원: 권기범

- 프론트엔드
- UI / UX 개발
- 실시간 API연동



팀원: 김현진

- 백엔드
- 데이터셋 수집, API 연동
- MySQL 구축

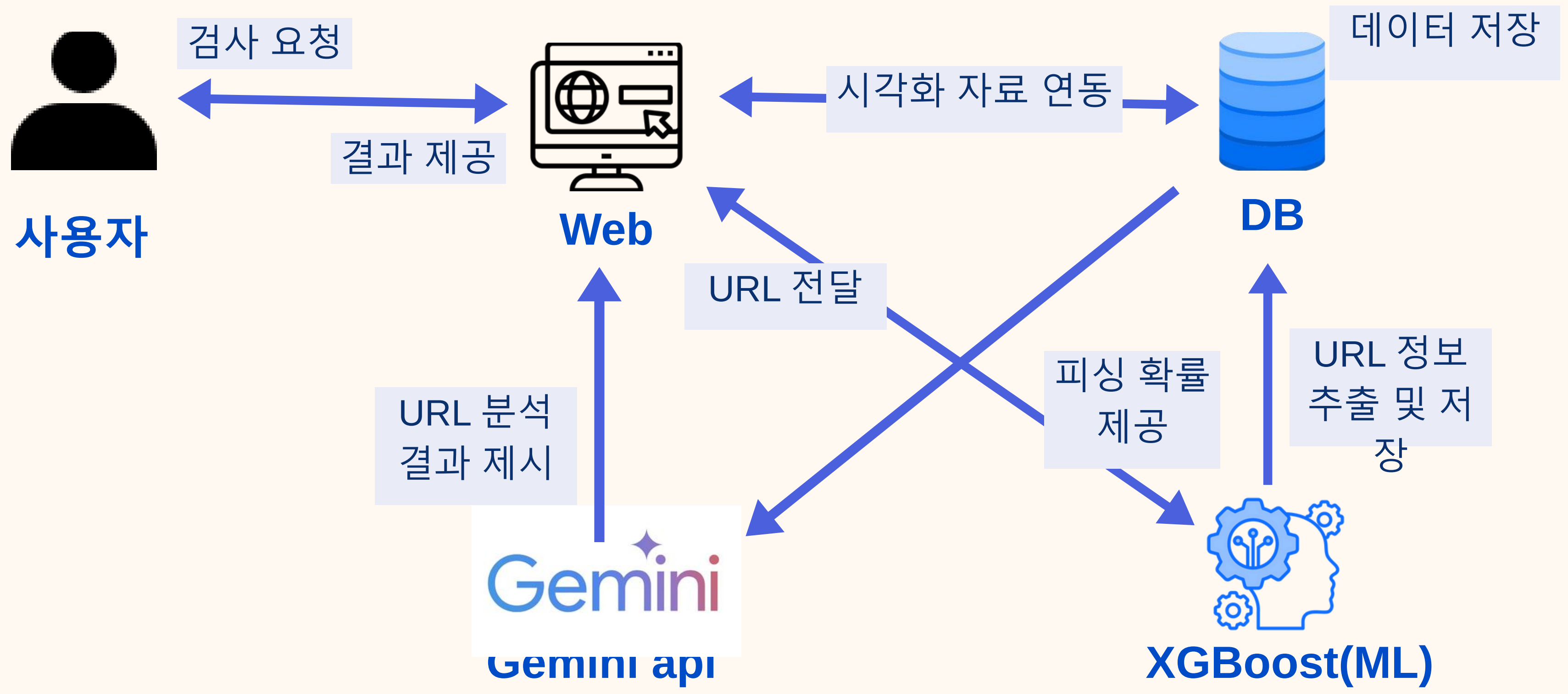
프로젝트 진행을 위해 세운 일정 계획입니다.

	3월	4월	5월	6월	7월	8월	9월	10월
기획 및 연구								
기본설계								
제작 및 구현								
테스트 및 오류 수정								
보고서 작성								

프로젝트 진행



프로젝트의 구조를 보기 쉽게 정리한 구성도입니다.



프로젝트를 진행하는 데에 있어서 사용한 개발 환경입니다.

백엔드 개발



- 머신러닝
- 데이터 처리
- API 제공



프론트엔드 개발



- 웹 디자인
- UI 스타일링

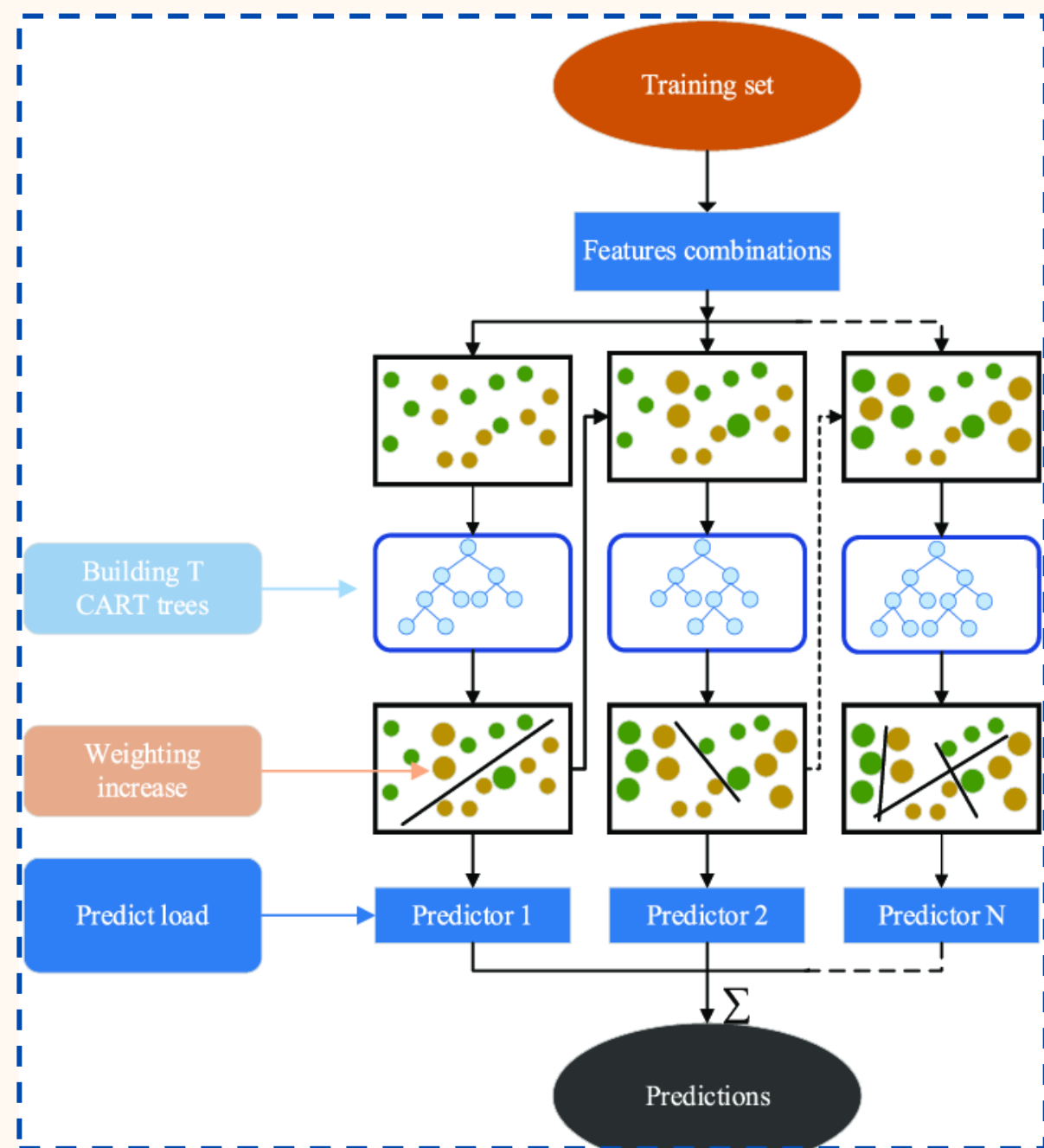


- 웹 페이지 구조



- API 연동

데이터셋을 학습하고 URL을 분석하는 머신러닝 모델로 'XGBoost'를 사용.



‘XGBoost’란?

- 병렬 연산과 캐시 최적화로 대용량 데이터셋도 효율적으로 학습.
- 모델 복잡도 제어를 통한 과적합 방지.
- 데이터 불완전성에도 안정적인 성능 유지.
- 각 피처가 예측에 미치는 영향을 수치화하여 해석 가능.

‘XGBoost’ 선정 이유

- 학습 효율성, 안정적인 성능, 피처 중요도 시각화 등 → 보안 서비스에 적합한 신뢰성.

‘XGBoost’ 하이퍼 파라미터

- learning_rate: 모델이 최적값에 수렴하는 속도 조절. = ‘ 0.1 ’
- max_depth: 트리의 최대 깊이 제한으로 과적합 방지. = ‘ 5 ’
- n_estimators: 생성할 트리의 개수, 정확도와 연산 효율성의 균형점. = ‘ 300 ’

‘XGBoost’가 학습할 때 사용하는 데이터의 속성값, 피처.

구분	주요 피처	설명
URL 구조 기반	• URL 길이 • 도메인 길이 • 서브 도메인 개수 • 하이픈 · 언더스코어 포함 여부	• URL 자체의 복잡도를 수치화하여 피싱 URL의 난독화 정도를 분석. • 피싱 URL은 일반적으로 비정상적으로 길거나 복잡한 구조.
보안 속성 기반	• HTTPS 존재 여부 • 파일 확장자 • Tiny URL 사용 여부 • DNS TTL	• 도메인과 연결된 보안 요소를 분석하여 URL의 신뢰도와 정상 사이트 여부를 판단. • 피싱 사이트는 종종 보안 인증서가 없거나 URL 단축 서비스를 악용.
통계적 특성 기반	• 특수문자 비율 • 숫자 비율 • 문자 엔트로피 • 연속 숫자/문자 길이	• 피싱 URL은 종종 무작위 문자열이나 비 정상적인 문자 패턴을 포함하는 경향. • 문자열 패턴의 통계적 특성을 분석.

피처 선정 이유

- 접근성 높은 속성 + 탐지 근거로 해석 가능한 속성 중심 선정

• WHOIS 기반 피처(도메인 생성일, 만료일)는 네트워크 질의 오류 고려하여 보조 피처로 분류
- 머신러닝 모델 입력을 위한 명확한 수치화 가능성 고려

• 결과 해석 및 설명 가능성을 확보하여 탐지 근거 제공

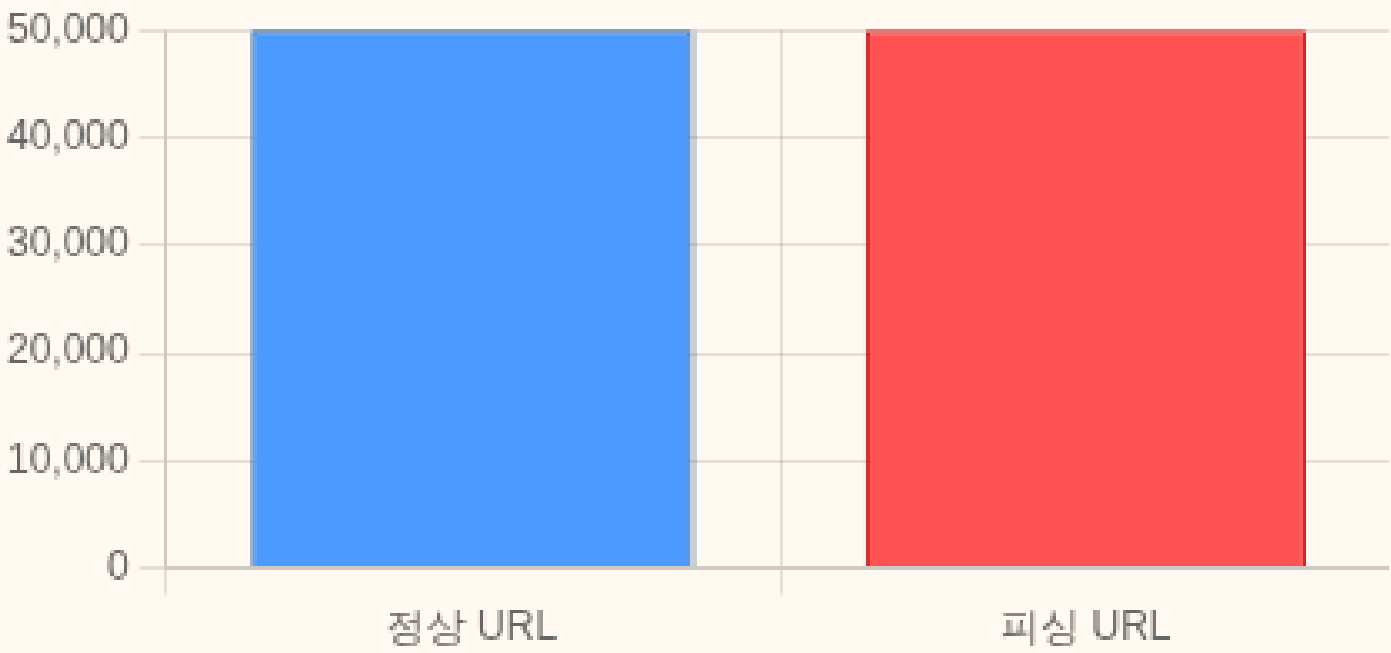
‘XGBoost’가 학습한 데이터셋 설명.

데이터셋 구성 및 전처리

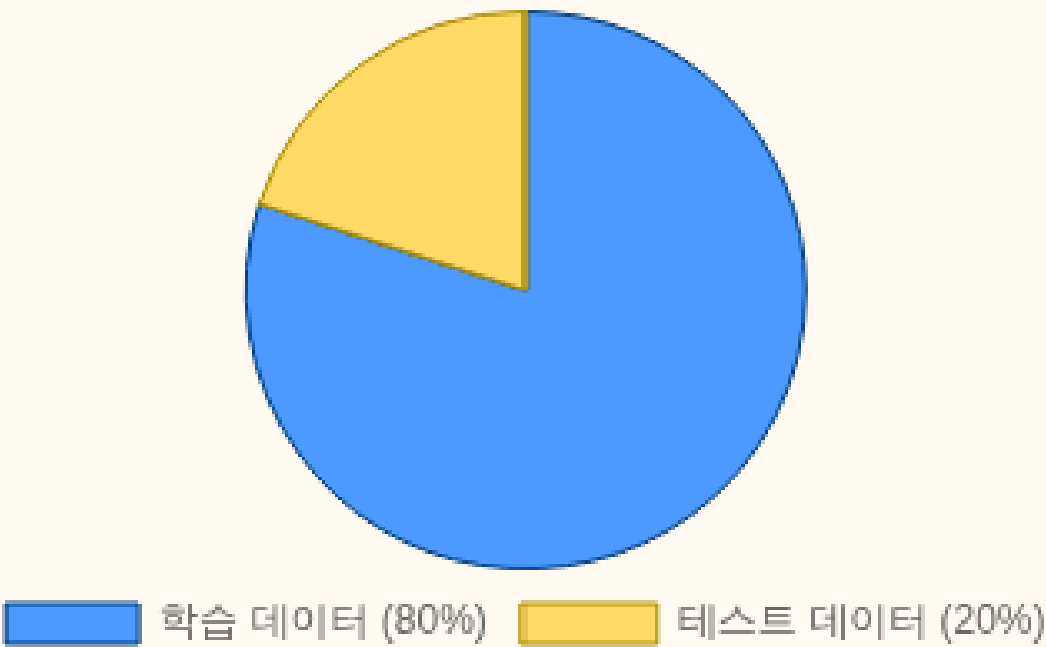
- Kaggle 공개 URL 데이터셋 활용
- 총 약 110,000개 URL (정상: 55,000개 / 피싱: 55,000개) 1 : 1 균형 데이터셋으로 정제 → 학습 : 테스트 비율 8 : 2 분할 (트레이닝 88,000개, 테스트 22,000개)



데이터셋 구성



학습/테스트 데이터 분할



‘XGBoost’의 탐지 성능 및 주요 피처에 대한 종합 분석.

탐지 모델 성능 및 주요 피처 분석

99.64%

정확도 (Accuracy)

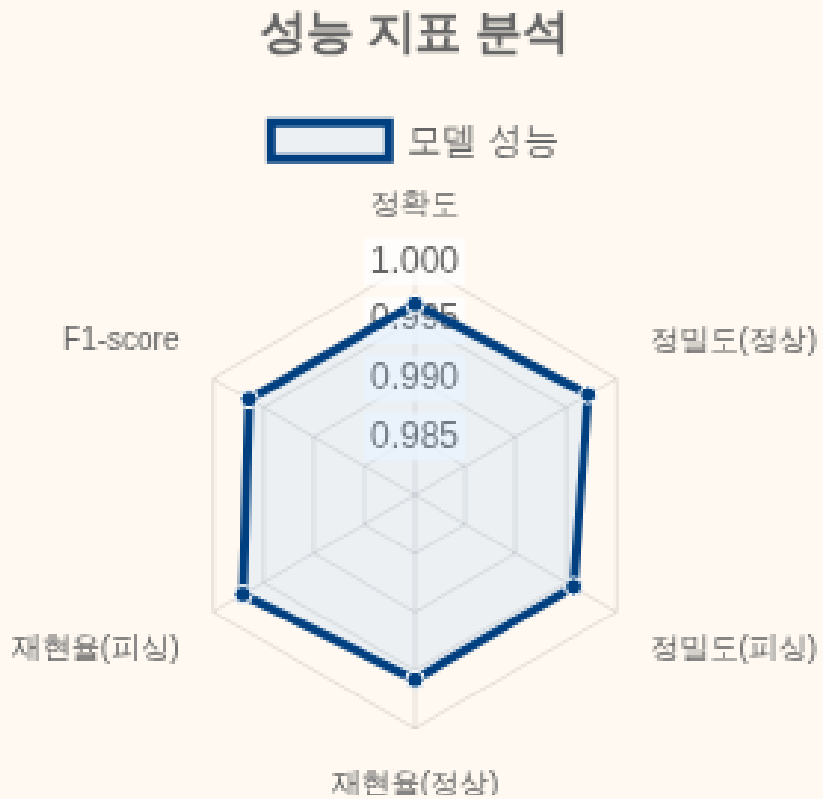
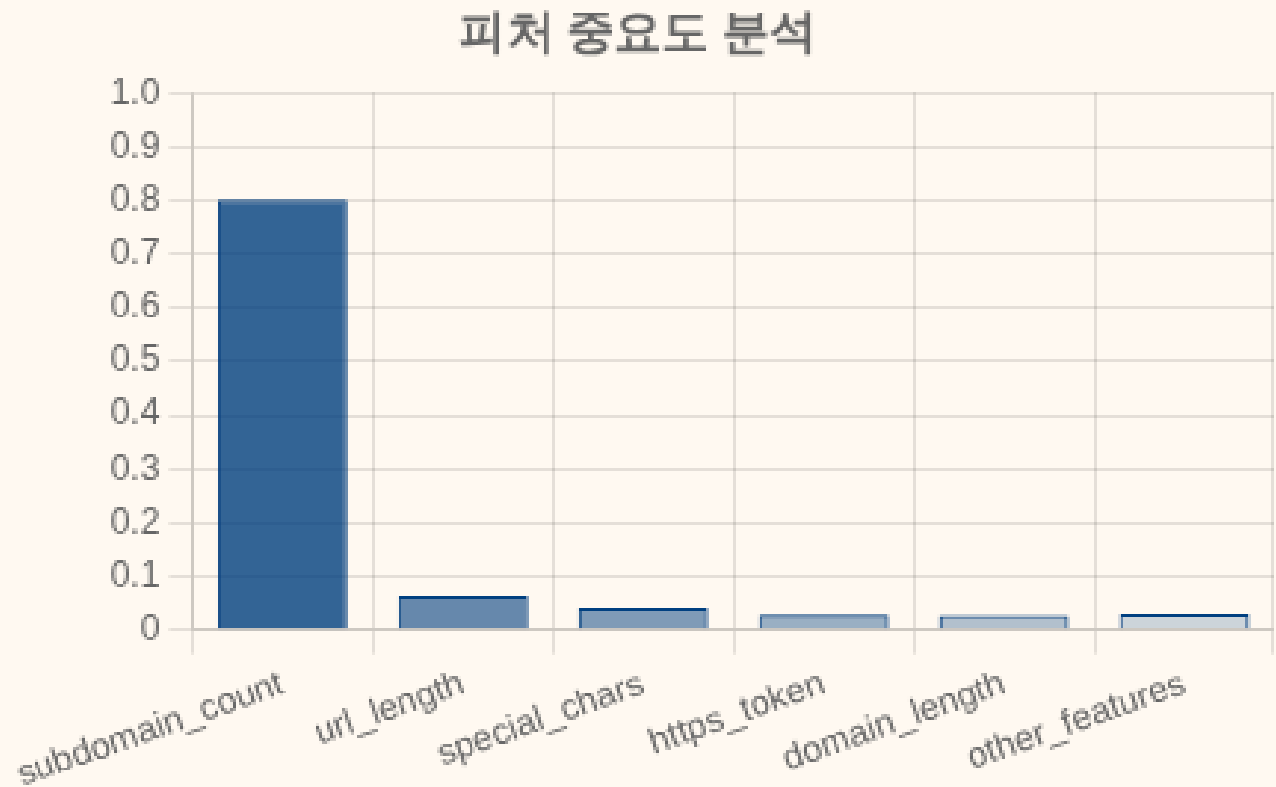
~1.00

정밀도/재현율
(Precision/Recall)

~1.00

F1-score
(정밀도/재현율의 조화 평균)

- 전체 데이터셋에 대한 탐지 정확도: 0.9964 (99.64%).
- 정상(0)과 피싱(1) 모두 정밀도, 재현율, F1-score가 1.00에 근접한 높은 성능.
- 가장 중요한 피처: subdomain_count (0.8019) - 피싱 URL은 서브도메인을 과도하게 포함하는 경향.



프로젝트 결과



AI를 활용한 URL 탐지 및 분석 웹 서비스 입력창 GUI.

피싱 사이트 탐지

한 번의 클릭으로 URL을 분석하고, 피싱 위험 여부를 확인하세요.

URL을 입력하세요 (예: https://example.com)

분석

딱걸렸조의 주요 기능

머신러닝 기반 탐지

XGBoost 모델을 이용한 높은 정확도의 피싱 탐지

WHOIS/DNS/IP 분석

도메인 생성일, DNS, IP 정보를 활용한 다층 보안 분석

실시간 시각화 맵

피싱 캠페인 활동을 세계 지도에서 실시간으로 확인

AI를 활용한 URL 탐지 및 분석 웹 서비스 정상 URL 결과 GUI.

피싱 사이트 탐지

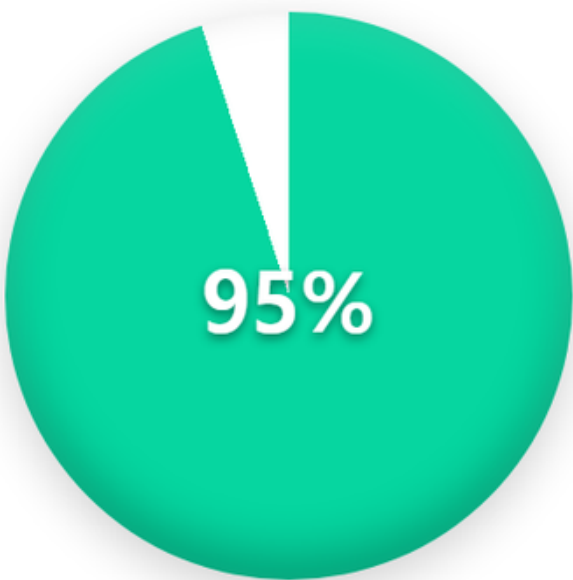
한 번의 클릭으로 URL을 분석하고, 피싱 위험 여부를 확인하세요.

https://www.naver.com

분석

분석 결과

정상일 확률 95.06%



분석 포인트



이 그래프는 상위 10개 요약입니다. 빨간 막대는 피싱 방향, 초록 막대는 정상 방향 기여도를 뜻합니다.

기록

그래프 상세보기

권장조치

▼ 🤖 AI 권장조치 — 정상 95.06%

모델이 이 URL을 정상으로 판단했습니다(정상 95.06%). 핵심 근거: 특수문자 비율 +31.78% — 특수문자 과다는 난독화/리다이렉션 유도 신호일 수 있습니다.; URL 길이 +24.86% — 주소가 과도하게 길면 피싱 가능성이 높아지는 경향이 있습니다.; 서브도메인 개수 +9.85% — 서브도메인이 많을수록 사칭/리다이렉트 구조일 수 있습니다.. 반대 신호도 관찰됨: 호스트 이름 길이

신고

신고방법 안내

처리절차

‘신고방법 안내’는 KISA 공식 가이드를, ‘처리절차’는 접수 후 진행 과정을 보여줍니다. ‘신고’는 실제 접수 폼으로 이동합니다.

AI를 활용한 URL 탐지 및 분석 웹 서비스 피싱 URL 결과 GUI.

피싱 사이트 탐지

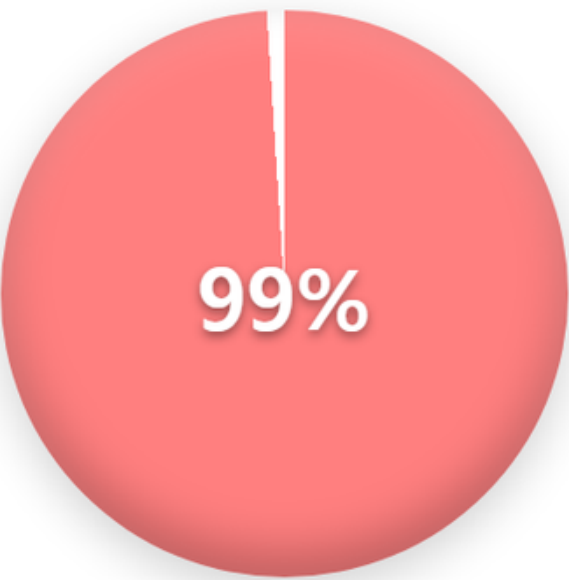
한 번의 클릭으로 URL을 분석하고, 피싱 위험 여부를 확인하세요.

http://limited-zoo-crooked.on-fleek.app/xnohzcyqmrp1.html

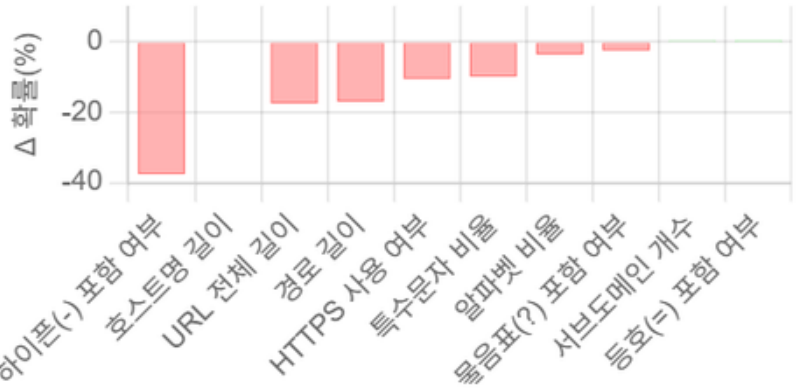
분석

분석 결과

피싱일 확률 99.17%



분석 포인트



이 그래프는 상위 10개 요약입니다. 빨간 막대는 피싱 방향, 초록 막대는 정상 방향 기여도를 뜻합니다.

기록

그래프 상세보기

권장조치

AI 권장조치 — 피싱 99.17%

모델이 이 URL을 피싱으로 판단했습니다(피싱 99.17%). 핵심 근거: 하이픈(-) 포함 여부 -37.61% — 브랜드 철자를 하이픈으로 쪼개 사칭하는 패턴이 흔합니다.; URL 길이 -17.69% — 주소가 과도하게 길면 피싱 가능성이 높아지는 경향이 있습니다.; 경로(path) 길이 -17.04% — 경로가 지나치게 길거나 복잡하면 리다이렉트/난독화 신호일 수 있습니다.. 반대 신호도 관찰됨: 호스트

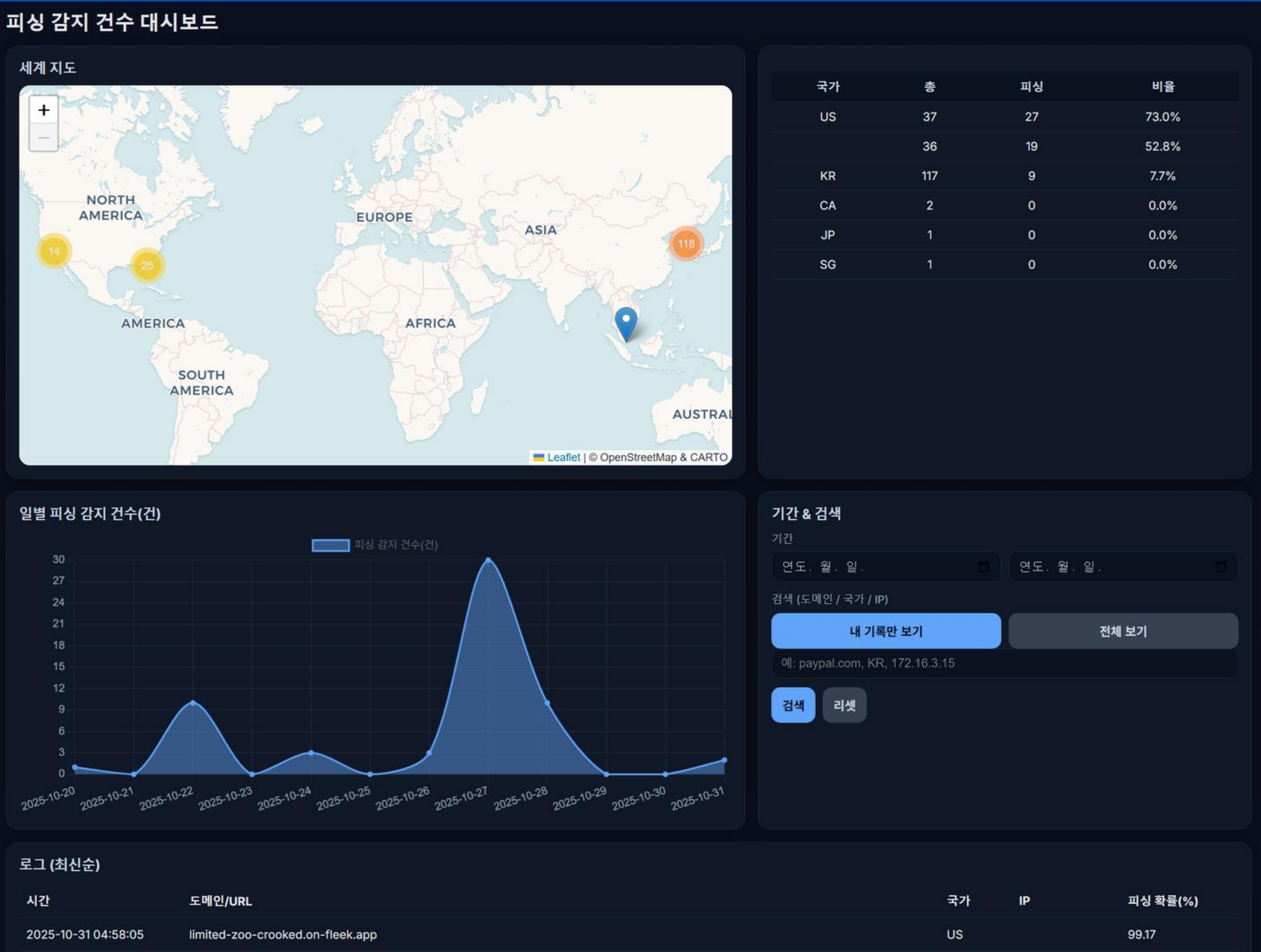
신고

신고방법 안내

처리절차

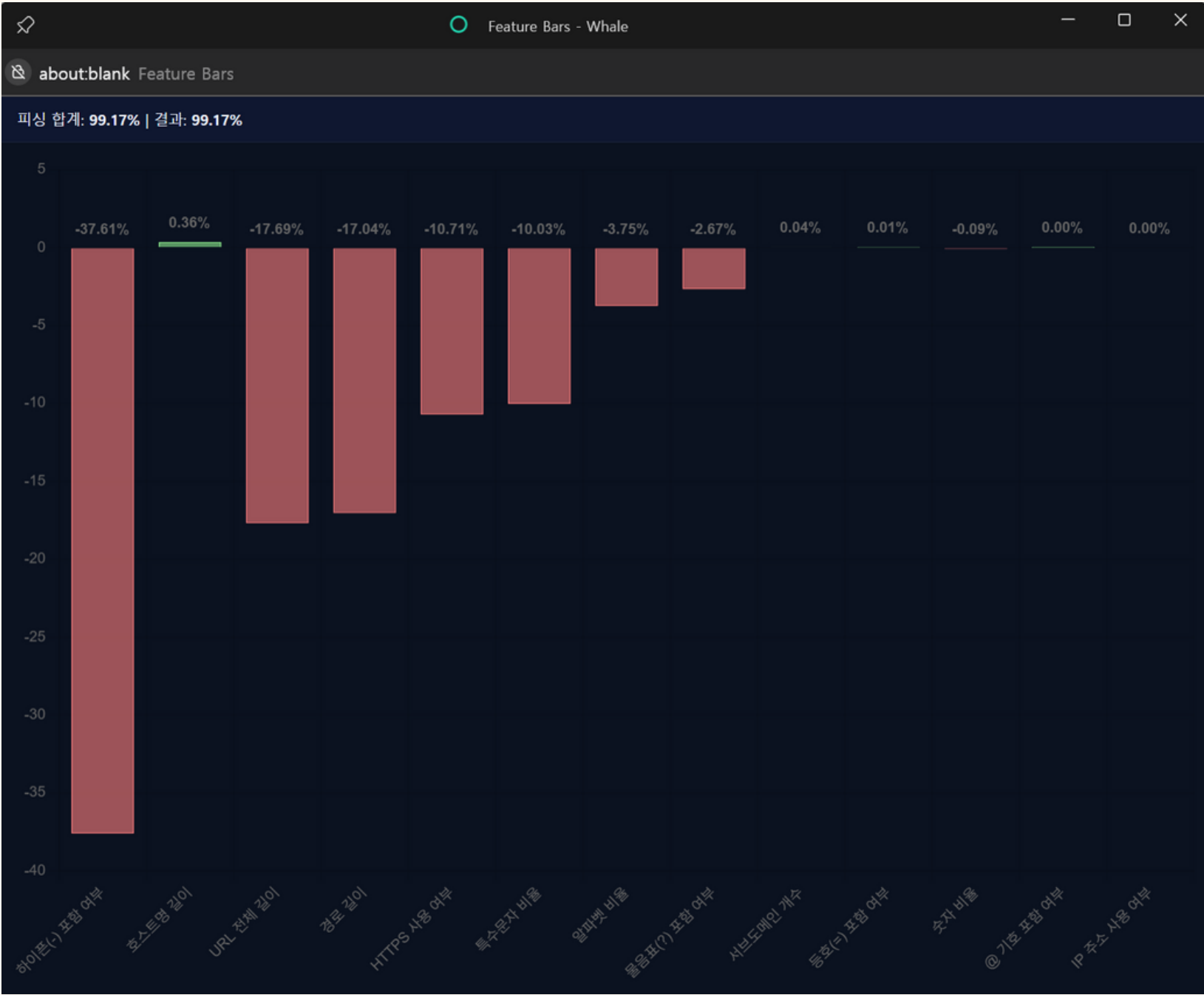
‘신고방법 안내’는 KISA 공식 가이드를, ‘처리절차’는 접수 후 진행 과정을 보여줍니다. ‘신고’는 실제 접수 폼으로 이동합니다.

DB에 저장된 내용을 시각화하여 보여주는 GUI.



- 피싱 URL의 데이터만 시각화.
- 입력된 피싱 URL 전부 확인 가능 및 사용자가 입력한 피싱 URL만 확인도 가능.
- 국가 정보를 통해 세계 지도에 표시
→ 국가별 피싱 비율 확인.

피처 기반 분석 포인트 상세 보기 및 AI의 분석과 권장 조치 GUI.



- 그래프 상세보기를 통해 열리는 팝업창.
- 피처 기반으로 분석 결과의 근거를 설명.
- 정상일 때와 피싱일 때 모두 확인 가능.
- ML이 분석할 때 중요하게 판단한 결과를 시각화.

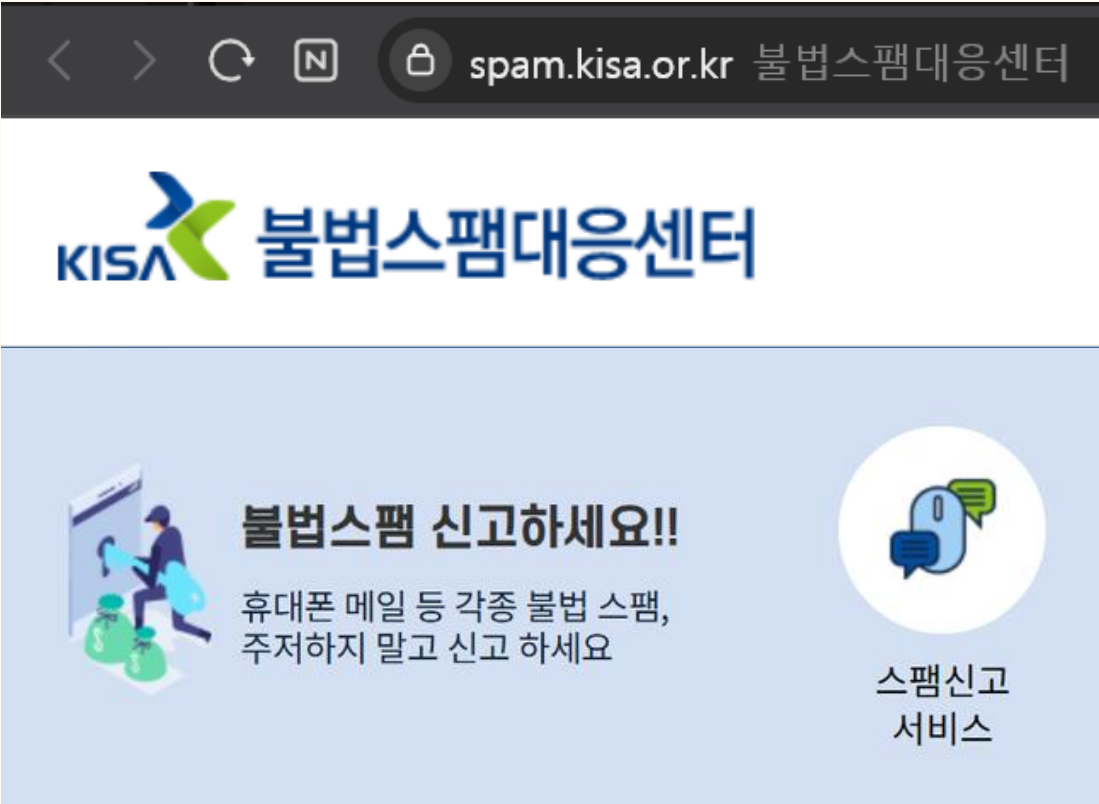
권장조치

▼ 🤖 AI 권장조치 — 피싱 99.17%

모델이 이 URL을 피싱으로 판단했습니다(피싱 99.17%). 핵심 근거: 하이픈(-) 포함 여부 -37.61% — 브랜드 철자를 하이픈으로 쪼개 사칭하는 패턴이 흔합니다.; URL 길이 -17.69% — 주소가 과도하게 길면 피싱 가능성이 높아지는 경향이 있습니다.; 경로(path) 길이 -17.04% — 경로가 지나치게 길거나 복잡하면 리다이렉트/난독화 신호일 수 있습니다.. 반대 신호도 관찰됨: 호스트

- AI 권장 조치 버튼으로 생성.
- 분석 포인트를 상세하게 설명 및 권장 조치 안내.

피싱 URL 신고 권장 기능으로 KISA 신고페이지와 연결.



- KISA의 신고페이지로 넘어가 신고 가능.
- KISA를 몰랐던 사람들에게 KISA 소개.

• 신고 방법과 처리 절차 안내.

스팸신고 방법

- ● **휴대전화 스팸신고 방법**
 - 1 **간편신고 서비스**
 - 휴대폰에서 간단한 버튼 조작만으로 스팸을 신고할 수 있는 무료서비스이며, 신고 시 한국인터넷진흥원 불법스팸대응센터로 신고가 됩니다.
 - 본 홈페이지의 '불법스팸 > 스팸신고 처리결과'에서 신고 내역 및 처리결과를 확인할 수 있습니다.
 - ※ 현재 iOS 및 일부 휴대전화 단말기에서는 간편신고 기능을 이용할 수 없습니다.
 - 2 **불법스팸 간편신고 앱**
 - 휴대전화 음성/문자 스팸을 간편하게 신고하고, 신고 내역 및 처리 결과를 확인할 수 있는 앱으로 iOS, 안드로이드 환경에서 모두 이용 가능합니다.
 - 설치 바로가기 →
- ● **이메일 스팸신고 방법**
 - 본 홈페이지의 '불법스팸 > 스팸신고 > 이메일 스팸신고' 탭에서 이메일 스팸신고를 접수 할 수 있습니다.
- ● **기타(팩스, 게시판 등)**
 - 본 홈페이지의 '불법스팸 > 스팸신고' 페이지를 통해 기타(팩스, 게시판 등) 스팸신고를 접수 할 수 있습니다.

스팸신고 접수 및 처리 절차

사용자 신고 단계

불법스팸신고

신고 접수, 위법사실 확인 단계

처리완료
수사의뢰
과태료 부과 의뢰

신고 처리 단계

- KISA(법위반사실 없음)
- 검, 경찰(수사)
- 방송통신사무소(수사/과태료)

프로젝트를 발전시키기 위한 추가 계획입니다.

데이터셋 확장 및 다양화

국내외 피싱 신고 데이터와 실시간 크롤링을 통해 다양한 URL 확보, 자동 수집 파이프라인 구축으로 최신 피싱 트렌드 반영

WHOIS/DNS 기반 피쳐 복원 및 개선

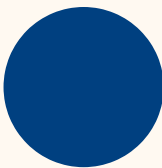
API 질의 방식으로 도메인 기반 피쳐 안정적 복원, 도메인 수명, 등록 기관, 네임서버 일치 여부 등 신뢰도 분석 기능 향상

웹 서비스 고도화 및 사용자 편의성 강화

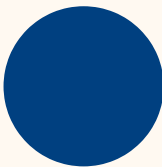
관리자용 대시보드 추가, 탐지 결과 통계, 주요 피쳐 기여도, 피싱 패턴 변화 시각화 기능 제공, 서비스 배포 및 운영 환경 구축

학문적 및 산업적 확장성

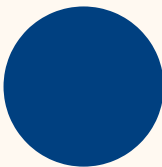
이메일·메신저 피싱 탐지 시스템으로 확장, 금융권 및 보안 기관의 사이버 위협 인텔리전스(CTI) 플랫폼 연계 가능성 검증



단기: 데이터셋 200만 URL 확장



중기: API 서비스화 및 플러그인 개발



장기: 통합 보안 플랫폼으로 발전

프로젝트 결과: 시연 영상

지금까지 발표한 프로젝트의 시연 영상입니다.

<https://youtu.be/4WYzcpJInpE>

$Q_n A$



이상으로 딱걸렸조의 졸업 작품 발표를 마치며 감사드립니다.

QnA