

「K-사이버 시큐리티 챌린지 2019」 설명회
[정보보호 R&D 데이터 챌린지]
- AI 기반 악성코드 탐지 트랙



이은지 주임연구원

한국인터넷진흥원 보안기술확산팀

2019. 10. 01

Contents

- 1 ▶ 트랙 개요
- 2 ▶ 활용 데이터셋
- 3 ▶ 결과물 제출 및 심사기준
- 4 ▶ 예선 진행방식
- 5 ▶ 본선 진행방식
- 6 ▶ 시상내역
- 7 ▶ 접수기간 및 신청절차



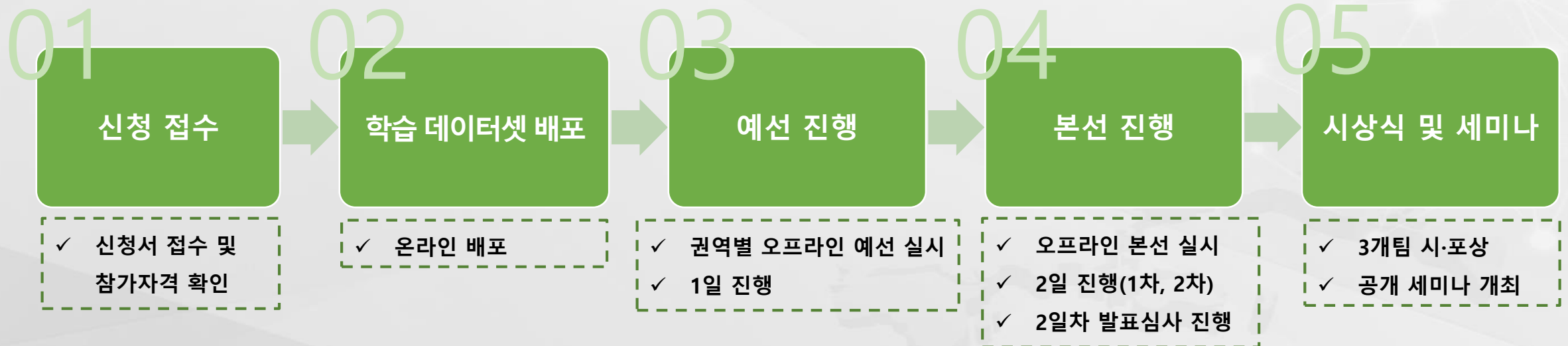
1. AI 기반 악성코드 탐지 트랙 개요

AI기반 악성코드 탐지 트랙

개요

- 참가자가 개발한 AI 기반 악성코드 탐지 모델을 통해 악성코드 탐지 성능 경쟁을 위한 정보보호 R&D 데이터 챌린지 2019 「AI 기반 악성코드 탐지」 트랙 개최

세부 절차



2. 활용 데이터셋

AI기반 악성코드 탐지 트랙

- (규모) 정상/악성코드 약 4만개
 - (수집방법) KISA, 안랩, 이스트시큐리티, 하우리, 세인트시큐리티 등 국내 백신사(社) 공동구축
 - (데이터셋 구성) 악성코드 분류 및 분석정보 기반 대회용 데이터셋 구성
- 오픈소스 기반 AI모델 활용, 자체 탐지율 테스트 및 결과에 따른 데이터셋 추가 가공

학습 데이터셋

- ✓ (규모) 정상/악성 코드 10,000개
- ✓ (구성) 데이터셋, 정답지(.csv)

예 · 본선 데이터셋

- ✓ (예선 규모) 정상/악성코드 10,000개
- ✓ (본선 규모) 본선 1차/2차 각각 정상/악성코드 10,000개 : 총 20,000개
- ✓ (구성) 데이터셋 (정답지 미포함)

2-1. 활용 데이터셋

■ 데이터셋 설명

- (동작환경) 32/64 bit Windows
- (파일명) MD5, 확장자 .vir
- (고려사항) Windows32/64 bit 파일 혼합 구성, 스크립트 파일, Anti-머신러닝, 패킹
 - Anti-머신러닝 : 강화학습 기반의 적대적 머신러닝 오픈소스 도구 활용·구성

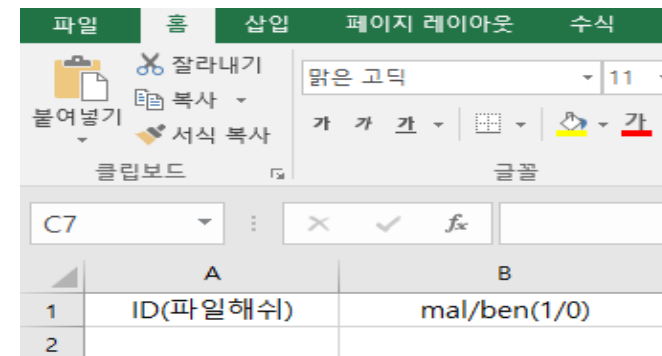
※ '19년도 데이터 가공 과정에서는 '18년 챌린지의 Installer형 Anti-머신러닝은 고려하지 않음

■ 정답 및 결과 파일

※ 파일 형식 : csv

※ 라벨 : 정상 0, 악성 1

※ 스크립트 정의 : hwp, html 등



	A	B
1	ID(파일해쉬)	mal/ben(1/0)
2		

3. 결과물 제출 및 심사기준

AI기반 악성코드 탐지 트랙

1. 결과파일

- ✓ 탐지결과를 파일명 ID와 정상/악성을 라벨로 분류하여 CSV파일 형태로 작성
- ✓ ID : MD5값, 라벨 : 정상코드 0, 악성코드 1

2. 알고리즘 설명문서

- ✓ **필수 내용** : 데이터 분석및분류 결과, 선정 feature 설명, 알고리즘 구성 방법
- ✓ **추가 내용** : 수도코드 (pseudocode), 예선 데이터 실험과정, 예상 결과, 보완점 등을 작성
- ❖ 추후 홈페이지(www.datachallenge.kr)를 통해 튜토리얼 공개
 - ※ 알고리즘 설명문서를 검토하여 편법 사용 등 문제 발견 시, 수상에서 제외될 수 있음
 - ※ 필요시 실험과정 영상/로그 자료 요청 예정 (준비 必)

3. 발표자료

- ✓ 알고리즘 설명문서 요약, 본선 데이터 분류 방법, 탐지 결과 등을 포함하여 **15분 분량**으로 작성
- ✓ **평가 내역** : 탐지 방안 (feature 등) 의 창의성, 알고리즘 완성도 등
 - ※ 알고리즘 설명문서는 평가 참고자료로 활용 됨

4. 기타 (채점관련)

- ✓ (발표) **분석도구 직접 개발** 여부 채점 **가중치 부여**
- ✓ (탐지결과) **과탐률** 및 **미탐률**에 상이한 채점 **가중치 부여**

3. 결과물 제출 및 심사기준

- ✓ (예선) 탐지율 100%
- ✓ (본선) 탐지율 80% + 발표점수 20%

$$\text{탐지율} = \text{정탐률} - \{(\text{과탐률} \times 0.6) + (\text{미탐률} \times 0.4)\}$$

정탐률

$$\frac{TP + TN}{TP + FP + FN + TN} \times 100$$

과탐률

$$\frac{FP}{FP + TN} \times 100$$

미탐률

$$\frac{FN}{FN + TP} \times 100$$

5. 채점 기준

카테고리		실제 결과	
		악성코드	정상코드
실험결과	악성코드	TP (True Positive)	FP (False Positive)
	정상코드	FN (False Negative)	TN (True Negative)

- ❖ True Positive : 실제 악성을 악성으로 정확히 예측
- ❖ True Negative : 실제 정상을 정상으로 정확히 예측
- ❖ False Positive : 실제 정상을 악성으로 예측 (과탐)
- ❖ False Negative : 실제 악성을 정상으로 예측 (미탐)

4. 예선 진행방식

AI기반 악성코드 탐지 트랙

I 예선 일정 및 진행방식

I (일정) 2019. 11. 9(토) 예정

I (구역) 4개 구역(서울·경기(강원)/충청(대전)/호남(광주,제주)/영남(부산,대구)) 오프라인 예선 개최

※ 참가 신청 시 구역 선택 必

※ 구역별 참가 신청자 3팀 이하인 경우 해당 구역 예선 취소

I (데이터셋 배포) 예선 대회 당일 오프라인 배포(USB)

I (점수산출 및 순위발표) 탐지율 100%를 반영하여 점수 산출 후, 최종 순위 발표

※ 편법여부 확인을 위해 알고리즘 설명 문서 제출

※ 본선진출팀 : 홈페이지 공지 및 개별 연락

I (본선 진출팀) 구역별 1,2위(총 8팀), 그 외 탐지율 상위 2개 팀 (총 10팀)

5. 본선 진행방식

I 본선 일정 및 진행방식

I (일정) 2019.11.21(목)~11.22(금)

※ 1일차 : 대회 개최 선언, 1차 데이터셋 분석, 결과파일 제출(1일차 종료 후 귀가)

※ 2일차 : 2차 데이터셋 분석, 결과파일 및 발표자료 제출, 발표 심사, 점수 산출 및 탐지율 순위발표

I (데이터셋 배포) 본선 대회 당일 오프라인 배포(USB)

I (결과파일 접수) 1일차, 2일차 각각 마지막에 제출한 결과를 최종 탐지 점수로 반영 - 1일차, 2일차 탐지 점수를 1:1로 반영(평균점수)

※ 대회 당일, 점수 보드를 통해 팀명-탐지율 송출

※ 제출 횟수는 홈페이지를 통해 추후 공지

I (발표심사 실시) 백신사(社) 관계자, 학계 전문가 등으로 구성된 평가위원회에서 심사 실시

※ 평가위원회 의견에 따라 추가 발표가 진행될 수 있음

I (점수산출 및 순위발표) 탐지율 80%, 발표 점수 20%를 반영하여 점수 산출 후, 최종 순위는 대회 홈페이지를 통해 발표



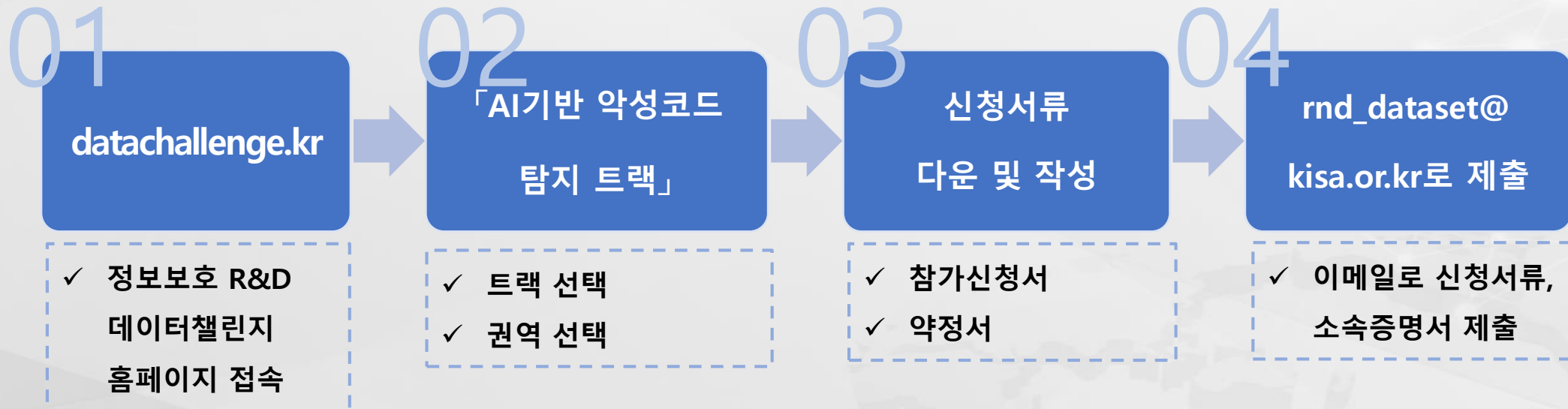
7. 접수기간 및 신청절차

AI기반 악성코드 탐지 트랙

접수기간

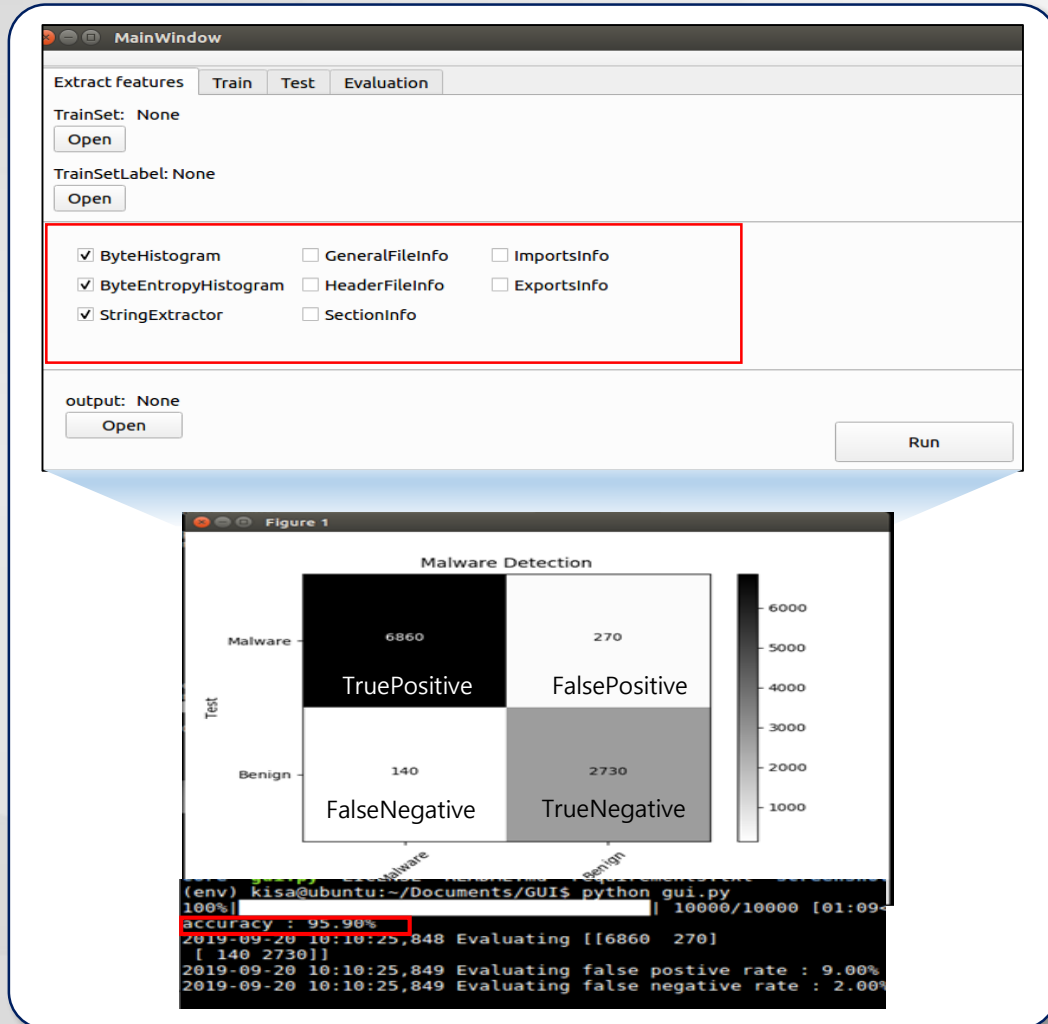
2019.10.1(화)~11.6(수)

신청절차



※ 권역별 오프라인 예선 개최를 위해 권역 선택 필수

- AI머신 학습지원을 위해 자체 개발한 『AI 기반 악성코드 탐지 시스템』 및 『'18 정보보호R&D 데이터챌린지』 데이터셋 제공(KISC 사이버보안빅데이터센터에 구축)



기능

- ✓ 사용자 지정·설정 feature를 기반으로 사용자 입력 데이터셋 학습

학습지원

- ✓ AI모델/데이터셋 대비 최적 feature 검증
- ✓ AI모델 성능 향상방안 및 저하원인 분석 등을 위한 학습지원

<신청/사용 문의>

E-mail) rnd_dataset@kisa.or.kr

Tel) 061-820-1323

Q&A



감사합니다.

<문의처>

한국인터넷진흥원 이은지 주임

(T. 061-820-1323)

(E. rnd_dataset@kisa.or.kr)