MACHINE LEARNING

네트워크 침입 탐지 시스템

이제우

삼성아카데미

예측 자동화 서비스과정

개인 프로젝트

목차

1. 개요

CIC-IDS-2017 기반 데이터를 이용하여 네트워크 트래픽 로그 기반으로 정상과 다양한 공격 유형을 자동으로 분류하는 네트워크 침입 탐지 시스템을 개발하여

침입탐지 성능을 측정하고, 주요 피처와 오류 원인을 분석해 실전 적용 가능성 평가

2. 계획

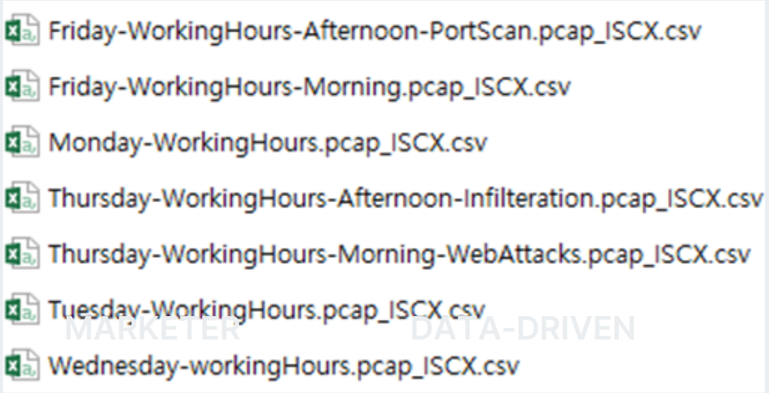
CIC-IDS-2017은 피처 기반 정형 데이터라 CNN/LSTM보단 MLP(DENSE 구조) 가 적합하다 판단, DENSE는 모든 피처 조합을 학습하기 때문에 희소 클래스 패턴까지 놓치지 않고 잡아낼 수 있을것으로 기대

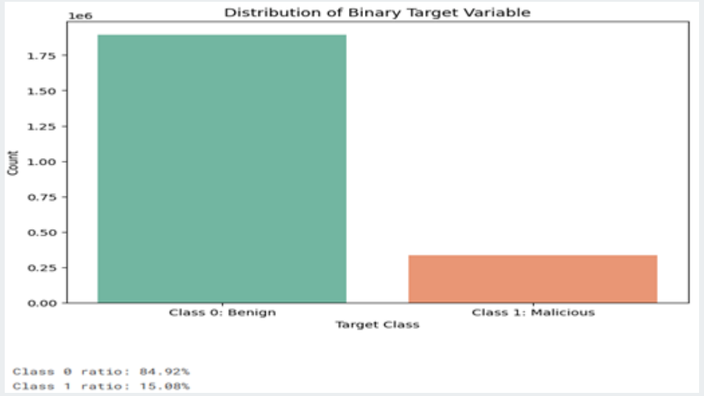
3. 데이터분석

캐나다 통신보안연구소(CSE) 산하 CIC에서 2017년7월3일~7일 5일간 수집하여 가공한 네트워크 공격 탐지 연구용 공개 데이터셋(IDS-2017)

8개의 CSV파일, 약 280,000행

문제: 정상 85%-공격15%의 불균형, 공격 라벨별 데이터 불균형





Label	
DoS Hulk	231073
PortScan	158930
DDoS	128027
DoS GoldenEye	10293
FTP-Patator	7938
SSH-Patator	5897
DoS slowloris	5796
DoS Slowhttptest	5499
Bot	1966
Web Attack ♦ Brute Force	1507
Web Attack ♦ XSS	652
Infiltration	36
Web Attack ♦ Sql Injection	21
Heartbleed	11

3. 데이터분석

70개의 피쳐중 중요 피쳐 분류결과:

BWD PACKET LENGTH MIN: 역방향(서버→클라이언트) 패킷 길이 최소값

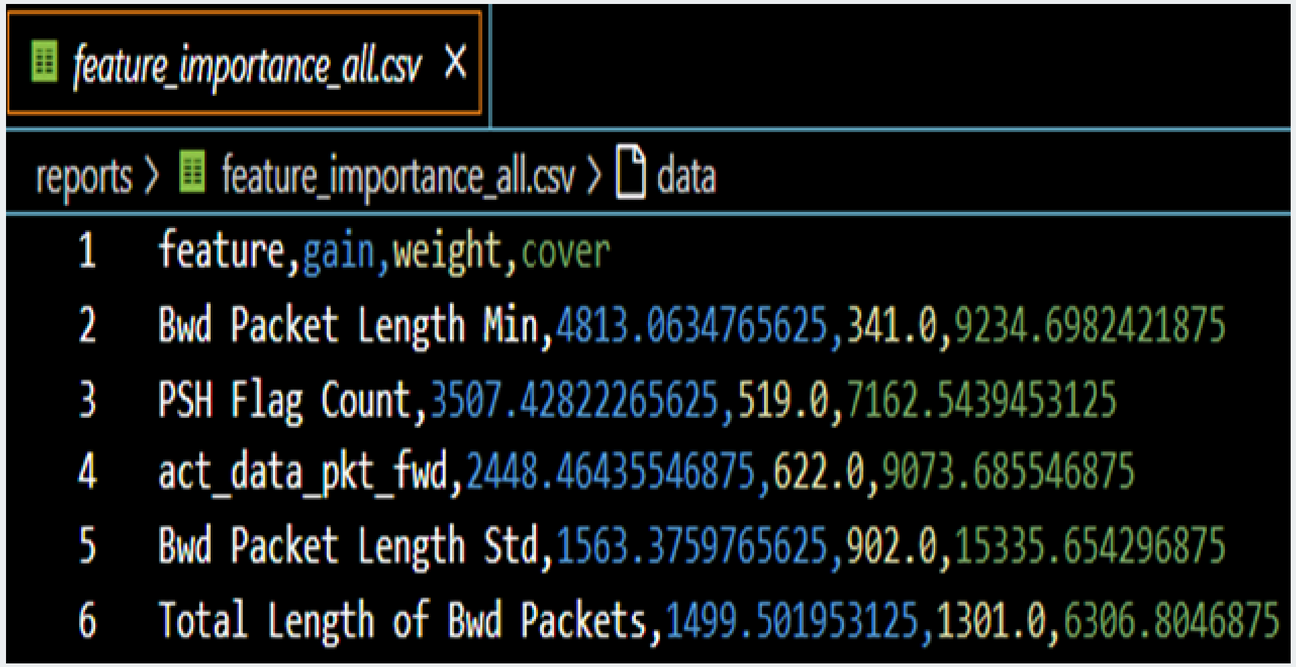
PSH FLAG COUNT: TCP PUSH 플래그가 설정된 패킷 개수

ACT_DATA_PKT_FWD: 순방향(ACTIVE) 데이터 패킷 수

BWD PACKET LENGTH STD: 역방향 패킷 길이의 표준편차

TOTAL LENGTH OF BWDPACKETS: 역방향 전체 바이트 합계

공격 트래픽은 보통 응답 패킷이 지나치게 짧거나 불규칙하고,
요청·응답 균형이 무너지며특정 플래그가 비정상적으로 많이발생하는것을 확인



4. 설계

1차BASELINE
(MLP, BATCH=512, EPOCHS=40, CLASS WEIGHT-100)

데이터 분할 TRAIN7:VAL1:TEST2

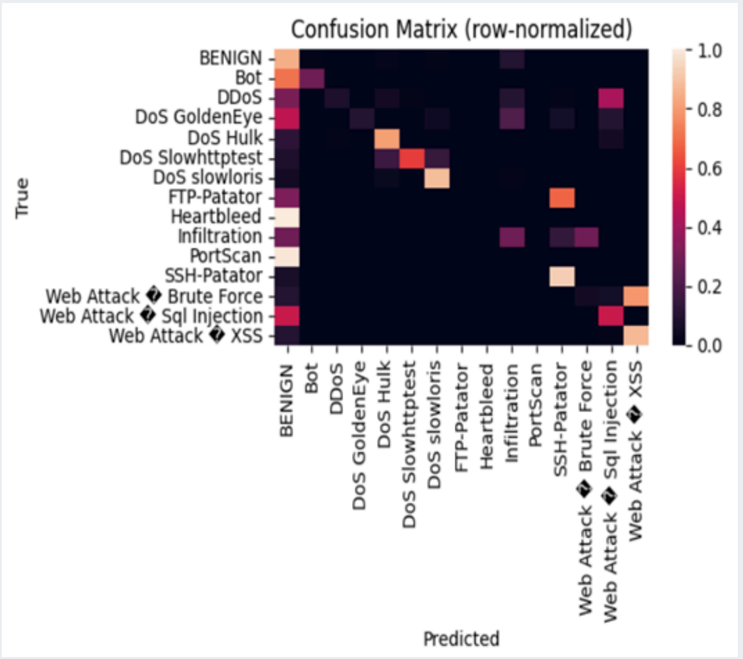
결과: BENIGN/대규모 공격만 잘 맞춤,
희소 클래스 RECALL↓

희소클래스들의 표본이 너무 적어 지표가 매우 불안정

극 희소클래스를 위한 과도한 가중치 편향으로인해
학습이 흔들렸다 판단

TEST-MACRO AVG: PRECISION 0.27 / RECALL 0.42 / F1 0.24

	precision	recall	f1-score	support
BENIGN	0.92	0.84	0.88	418276
Bot	0.03	0.29	0.05	391
DDoS	0.55	0.08	0.14	25603
DoS GoldenEye	0.82	0.09	0.17	2057
DoS Hulk	0.73	0.81	0.77	34570
DoS Slowhttptest	0.29	0.60	0.39	1046
DoS slowloris	0.12	0.89	0.21	1077
FTP-Patator	0.00	0.00	0.00	1187
Heartbleed	0.00	0.00	0.00	2
Infiltration	0.00	0.29	0.00	7
PortScan	0.00	0.00	0.00	18164
SSH-Patator	0.25	0.93	0.39	644
Web Attack &#65533; Brute Force	0.05	0.05	0.05	294
Web Attack &#65533; Sql Injection	0.00	0.50	0.00	4
Web Attack &#65533; XSS	0.32	0.88	0.47	130
accuracy			0.77	503452
macro avg	0.27	0.42	0.24	503452
weighted avg	0.84	0.77	0.79	503452



4. 설계

2차- 가중치 조정으로 보완 시도
(BENIGN축소,CAP=100, CLASS WEIGHT-70)

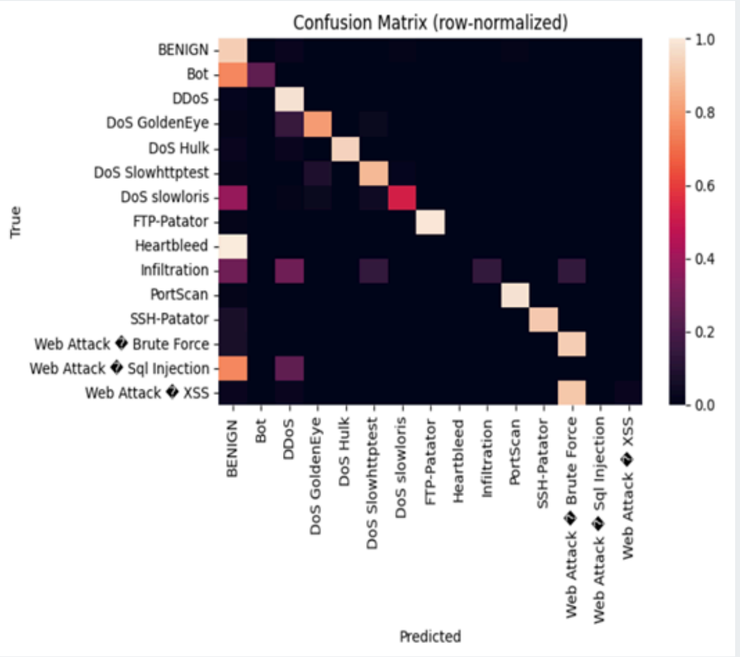
소수 클래스 RECALL↑

3차- 소수클래스 데이터 증강으로 보완 시도
SMOTE(CAP=15,000(최대 2배), CLASS WEIGHT-50))

소수 클래스 학습 기회 확대, 일부 F1-SCORE 개선

여전히 극심한 라벨간 불균형과 WEIGH의 편향으로 LOSS피크 발생

	precision	recall	f1-score	support
BENIGN	0.98	0.93	0.96	170351
Bot	0.45	0.25	0.32	391
DDoS	0.80	0.98	0.88	25603
DoS GoldenEye	0.69	0.80	0.74	2057
DoS Hulk	0.99	0.94	0.97	34570
DoS Slowhttptest	0.38	0.87	0.52	1046
DoS slowloris	0.19	0.53	0.28	1077
FTP-Patator	0.98	0.99	0.99	1187
Heartbleed	0.00	0.00	0.00	2
Infiltration	0.00	0.14	0.01	7
PortScan	0.89	0.98	0.93	18164
SSH-Patator	0.99	0.92	0.95	644
Web Attack Brute Force	0.67	0.93	0.78	294
Web Attack Sql Injection	0.00	0.00	0.00	4
Web Attack XSS	0.44	0.03	0.06	130
accuracy			0.94	255527
macro avg	0.56	0.62	0.56	255527
weighted avg	0.95	0.94	0.94	255527



4차(MLP최종)- 파라미터 튜닝 배치사이즈 조정
BATCH=512 → 256, CLASS WEIGHT-30

LOSS안정화, 성능 일부 개선
지표는 1차에 비해 다소 안정됐지만 희소클래스가 상대적 학습기회를 보장받지 못했음

최종 TEST-MACRO AVG: PRECISION 0.56 / RECALL 0.62 / F1 0.56

1차 대비 TEST-MACRO AVG: PRECISION 0.29 / RECALL 0.20 / F1 0.30 개선

4. 설계

CIC-IDS-2017은 피처 기반 정형 데이터로, 초기에는 MLP(DENSE 구조)가 적합하다고 판단

DENSE는 모든 피처 조합을 학습할 수 있어 희소 클래스 패턴까지 잡아낼 수 있을 것으로 기대하여, CLASS WEIGHT 조정과 배치 사이즈 변경을 통해 불균형 문제를 완화하고자 하였다.

또한 SMOTE를 적용해 희소 클래스 학습 기회를 확장하려 했으나, 시간상 실사용이 가능한 정도의 성능 개선에는 한계가 있다고 판단

이에 따라 최종적으로는 소수 클래스에 강하고 빠른 학습이 가능한 머신러닝 알고리즘인 XGBOOST로 방향 전환

4. 설계

트리·부스팅 모델(XGBOOST)은 숫자 피처에 최적화

소수 클래스가 TRAIN에 0개면 그 클래스는 못 맞춤 → 최소 샘플 보장으로 극소수 클래스 보완

데이터 분할 TRAIN7:VAL1:TEST2

하이퍼파라미터:
MAX_DEPTH=8 표현력 VS 과적합 균형

SUBSAMPLE/COLSAMPLE_BYTREE=0.8: 매 라운드 부분 샘플링으로 과적합 억제

VALIDATION SET을 이용해 EARLY STOPPING 및 하이퍼파라미터 튜닝

클래스 가중치(CLASS_WEIGHT =INV/INVSQRT): 정상-공격 불균형을 학습 단계에서 보정(소수 클래스 가중↑)

데이터 불균형 → (DATASET) 최소 학습샘플 보장 + (TRAIN) 가중치 학습 지원.

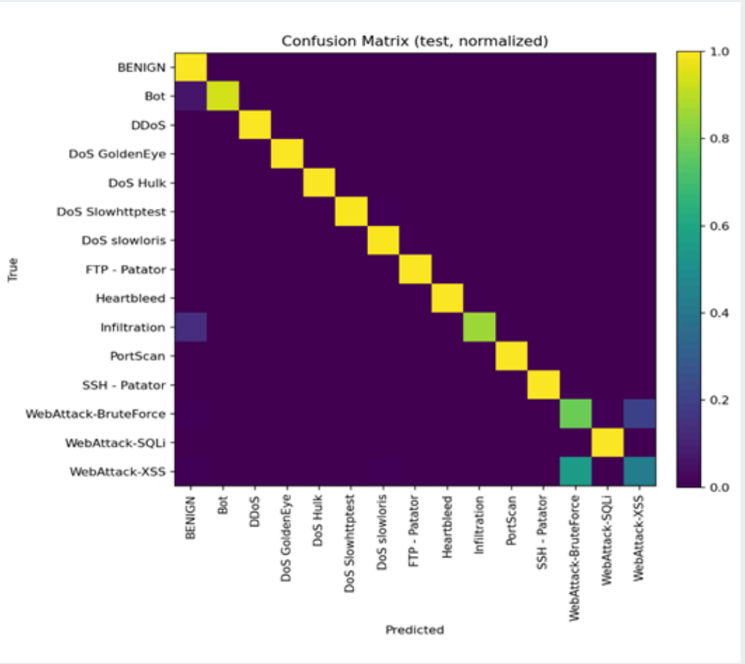
피처 누락/불일치 가능성 → (EVALUATE) ALIGN_TO_MODEL_FEATURES로 자동 보정

4. 설계

전체적으로 높은 점수를 기록하였으나 일부 희소 클래스의 경우
표본의 부족으로인해 점수가 다소 떨어지는 경향을 보였다.
특히 WEB ATTACK 류의 공격은
CIC-IDS-2017 같은 정형(FLOW)데이터 이기에
콘텐츠(문자열) 기반 WEB 공격은 구분하기 어려운 특징을 보였다.

reports > classification_report_test.txt

	precision	recall	f1-score	support
BENIGN	0.9999	0.9992	0.9996	454620
Bot	0.7764	0.9364	0.8489	393
DDoS	0.9998	0.9999	0.9999	25606
DoS GoldenEye	0.9985	0.9985	0.9985	2059
DoS Hulk	0.9986	0.9998	0.9992	46215
DoS Slowhttptest	0.9927	0.9936	0.9932	1100
DoS slowloris	0.9914	0.9957	0.9935	1159
FTP - Patator	1.0000	0.9994	0.9997	1588
Heartbleed	1.0000	1.0000	1.0000	2
Infiltration	1.0000	0.8571	0.9231	7
PortScan	0.9941	0.9996	0.9968	31786
SSH - Patator	0.9992	1.0000	0.9996	1179
WebAttack-BruteForce	0.7630	0.7807	0.7718	301
WebAttack-SQLi	0.6667	1.0000	0.8000	4
WebAttack-XSS	0.4508	0.4231	0.4365	130
accuracy			0.9990	566149
macro avg	0.9087	0.9322	0.9173	566149
weighted avg	0.9990	0.9990	0.9990	566149



VALIDATION-MACRO AVG: PRECISION 0.8586 / RECALL 0.8148 / F1 0.8284

TEST-MACRO AVG: PRECISION 0.9087 / RECALL 0.9322 / F1 0.9173

5. 평가

딥러닝MLP TEST-MACRO AVG: PRECISION 0.56 / RECALL 0.62 / F1 0.56

머신러닝XGBOOST TEST-MACRO AVG: PRECISION 0.90 / RECALL 0.93 / F1 0.91

정형 데이터와 불균형 라벨 분포에서는 복잡한 딥러닝 구조보다 전통적인 머신러닝 알고리즘이 더 효율적일 수 있다.

6. 향후계획

-앙상블 확장

파일별(PER-FILE) MLP 모델과 글로벌 모델을 조합하는 앙상블을 적용하여 단일 모델 대비 안정적이고 균형 잡힌 탐지 성능 확보

-시각화

FLASK 등을 활용하여 웹에서 동작할수 있는 구조로 확장

7. 출처

CIC-IDS-2017데이터-KAGGLE(CHETHANH N)

HTTPS://WWW.KAGGLE.COM/DATASETS/CHETHUHN/NETWORK-INTRUSION-DATASET/DATA

7. 출처

-END-