LID-DS Report 11/16

현재 진행상황

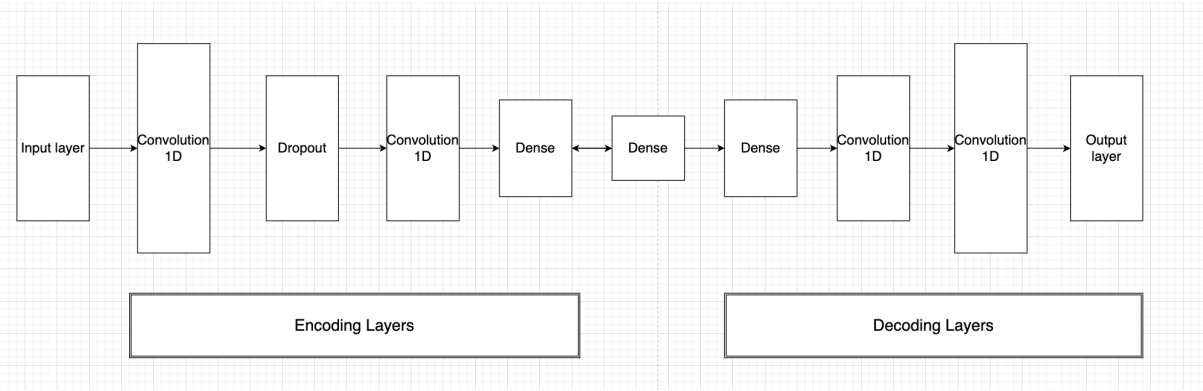
‘Bruteforce_CWE-307’, ‘CVE-2014-0160’ 두가지 공격을 단일모델로 학습하여 F1 score를 공격별 대략적으로 0.8의 성능을 뽑아내었다.

Test Data를 이용하여 성능평가를 할 때 시간소요로 인해 정상데이터만을 가지고 있는 시나리오는 300개, 공격을 포함하고 있는 시나리오는 100개를 선택하여 성능평가를 진행하였다.

Attack type	Train Data	Validation Data	Test Data(Only Normal)	Test Data(Contain Attack)
Bruteforce_CWE-307	40	10	300	100
CVE-2014-0160	40	10	300	100

모델 구성

Conv1d-Dense Denoising Autoencoder

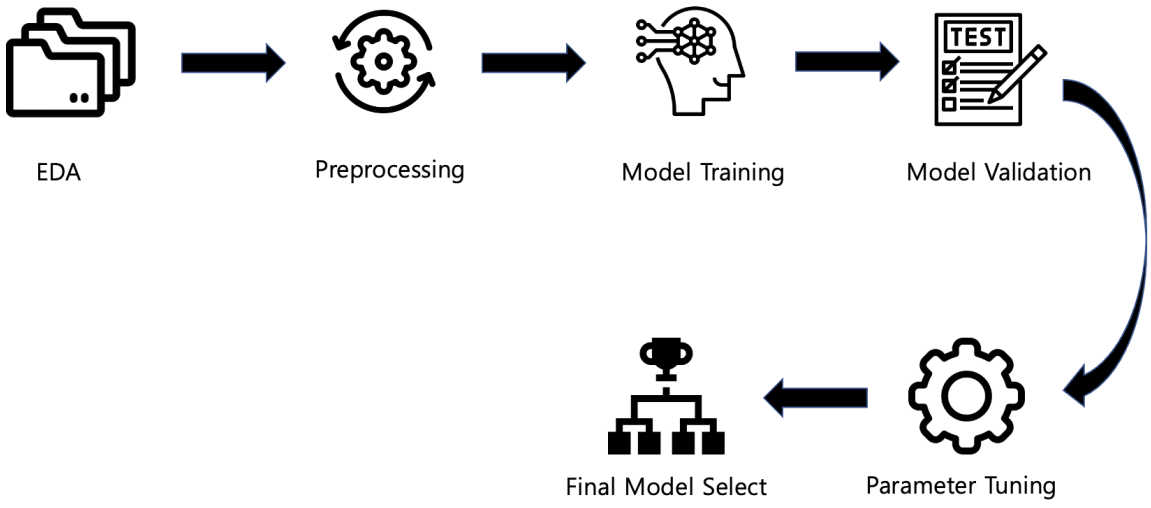


모델 파라미터 및 주요 속성

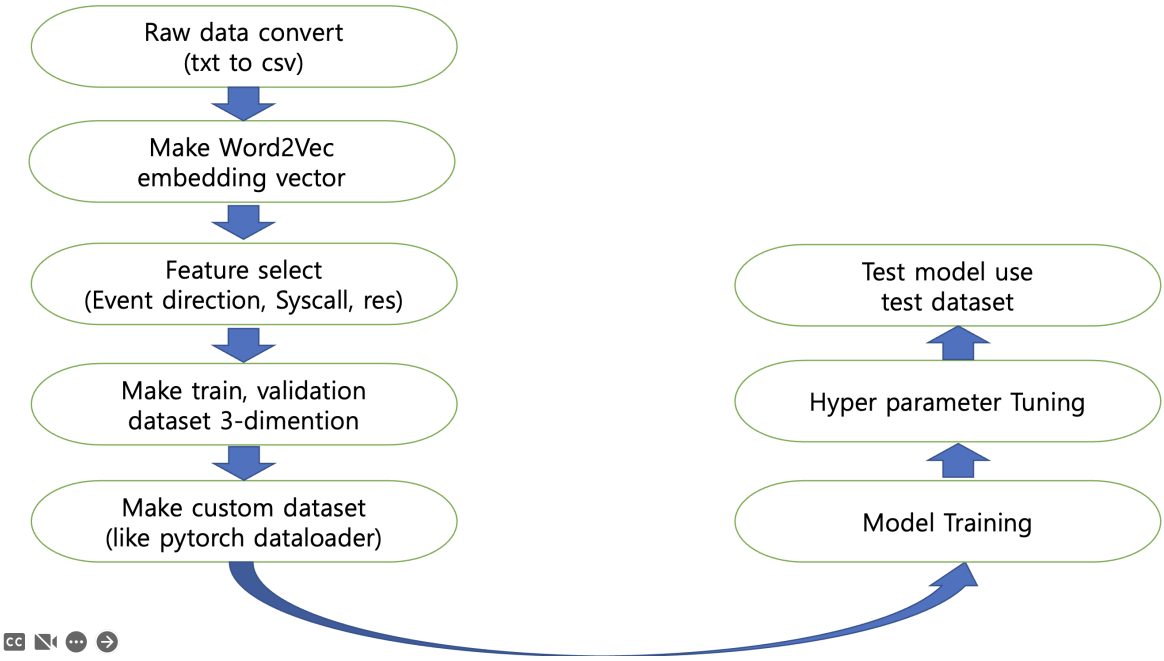
- Batch size = 32
- Epoch = 30
- Conv1d Layer kernel size = 3
- Optimaizer = adam
- Initial learning rate = 1e-3

- Kernel initializer = He initializer
- Activation function = relu
- Loss function = MSE
- EarlyStopping
  - patience = 5
  - monitor = val_loss
- Learning rate scheduler = ReduceOnLRPlateau(patience = 5)
- ModelCheckPoint
  - monitor = val_loss

대략적 학습 구성도



상세 학습 구성도



진행 예정 사항

- Keras의 Functional API를 이용하여 latent vector를 추출 후 t-SNE를 이용하여 latent vector 시각화
- ‘CVE-2012-2122’ 공격 데이터를 포함하여 총 3가지 공격에 대해 이상탐지를 진행할 예정
- 최종 목표는 ‘Bruteforce_CWE-307’, ‘CVE-2014-0160’ ‘CVE-2012-2122’ 3가지 공격 유형에 대해 F1 score 0.8 이상으로 이상을 탐지해내는 것이 이번 프로젝트의 목표이다.