

학습 세팅

- 모델
 - Flair 에서 제공하는 ner tagger model을 fine-tuning 함 (<https://huggingface.co/flair/ner-english>)
 - Flair embeddings 와 LSTM-CRF을 사용한 모델
- dataset split
 - corpus에 중복된 문장이 몇 개 포함되어 있었음 (51쌍). 해당 문장의 중복을 제거
 - 두개 이상의 랜섬웨어 이름이 들어간 문장을 제외. (제외 후 6540문장)
 - train, dev, test을 약 4:1:1의 비율로 나눔
 - 따라서 4360 train + 1090 dev + 1090 test sentences 으로 구성되었음.

랜섬웨어 이름 그룹핑

- train, dev, test 사이에 같은 이름의 랜섬웨어가 포함되지 않도록 그룹핑하여 배치함
- 그룹핑에 사용한 정규표현식
(TROJAN|RANSOMWARE|RANSOME|RANSOM|ANDROIDOS|ANDROID|VIRUS|_|-.\{1,2\}\$|\.\.\{1,2\}\$|V\d|\W|\d|S\$)
- 모든 글자를 대문자화 한 뒤, 위 정규표현식에 해당하는 문자들을 제거하여 랜섬웨어의 대표 이름만 남게 함
- 정규표현식 설명
 1. TROJAN|RANSOMWARE|RANSOME|RANSOM|ANDROIDOS|ANDROID|VIRUS : 랜섬웨어 이름에 포함되는 일반 명사를 제거
 2. _|\W|\d : 언더바, 공백, 숫자 등을 제거
 3. -.\{1,2\}\$|\.\.\{1,2\}\$|V\d : "-(버전)", ".(버전)", "V(버전)" 등의 랜섬웨어 버전 형식을 제거
 4. S\$: 복수형으로 단어 끝에 붙은 S를 제거

랜섬웨어 이름 그룹핑

- 그룹 목록
- 빈 문자열은 “RANSOM32”, “RANSOME32” 라는 이름의 랜섬웨어가 포함된 그룹이다.
- [' ', 'ALFA', 'ALPHA', 'AURORA', 'BADRABBIT', 'BITPAYMER', 'BUCBI', 'CERBER', 'CHACHA', 'CHIMERA', 'CLOP', 'CORONA', 'COVERTON', 'COVIDLOCK', 'CRYPTOBIT', 'CRYPTODEFENSE', 'CRYPTOHITMAN', 'CRYPTOJOKER', 'CRYPTOLOCKER', 'CRYPTOMIX', 'CRYPTOWALL', 'CRYPTXXX', 'CTBFAKER', 'CUTE', 'DJVU', 'DMA', 'DMACRYPTOLOCKER', 'DMALOCKER', 'EN', 'FUSOB', 'GANDCRAB', 'HDDCRYPTOR', 'HERMES', 'HYDRACRYPT', 'INVISIBLEEMPIRE', '**JIGSAW**', '**JIGSAWDAT**', 'KERANGER', 'KIMCILWARE', 'KOLER', 'KOLERANDSIMPLOCKER', 'KOLERPOLICELOCKER', 'KOVTER', 'LECHIFFRE', 'LECHIFFREE', 'LIME', 'LOCK', 'LOCKERPIN', 'LOCKY', 'MAGIC', 'MAKTUB', 'MAKTUBLOCKER', 'MANAMECRYPT', 'MAZE', 'MISCHA', 'MSILMANAMECRYPT', 'MYLITTLE', 'NANOLOCKER', 'NEFILIM', 'NEFILIMNEPHILIM', 'NEMTY', 'NOTPETYA', 'NYETYA', 'OSKOLER', 'OSXKERANGER', 'PADCRYPT', '**PETYA**', '**PETYAMISCHA**', 'POWERWARE', 'PROLOCK', 'RANSCAM', 'REVETON', 'ROKKU', 'RYUK', 'SAMSAM', 'SHADE', 'SIMPLEENCODER', 'SIMPLELOCKER', 'SIMPLOCKER', 'SODINOKIBI', 'STARE', 'STOP', 'STOPPROMORAD', 'TESLACRYPT', 'TORRENTLOCKER', 'TROLDESH', 'TRUECRYPT', 'TRUECRYPTER', 'UMBRECRYPT', 'UNIVERSALSYSTEMCRYPTOLOCKER', 'URAUSY', 'URAUSYPOLICE', 'WANNACRY', 'WINFILECODERHYDRACRYPT', '**WINLOCK**', '**WINLOCKDESKTOP**', '**WINLOCKER**', '**ZCRYPT**', '**ZCRYPTOR**', 'ZEMBLAXJIGSAW']
- 비슷한 이름이지만 다른 경우가 있어서 애매하다. (굵은 글씨)

랜섬웨어 이름 그룹핑

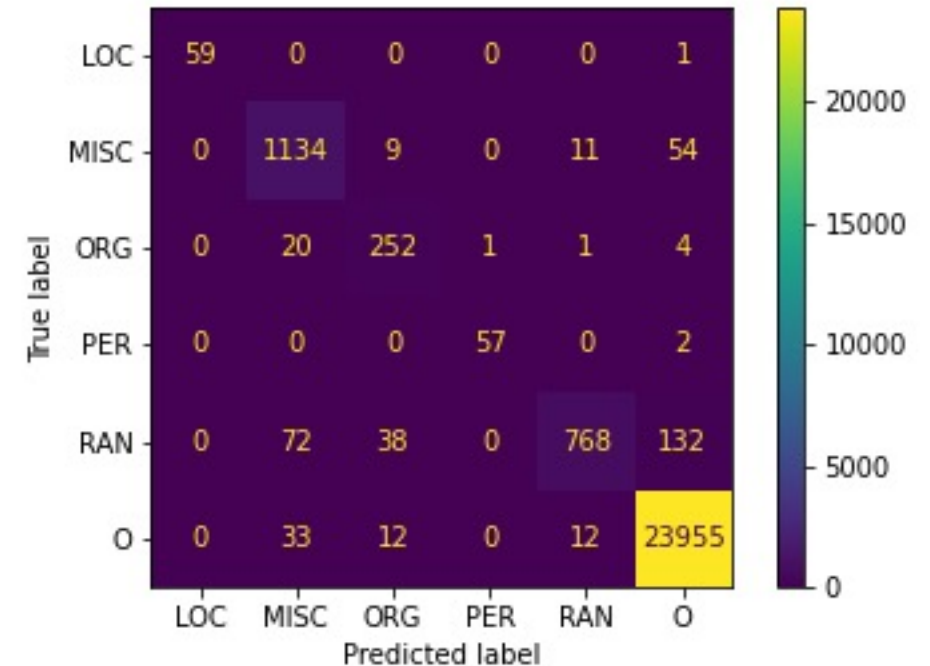
- 그룹핑에 정규 표현식을 사용하면 새로운 이름에 대해 예외가 생길 수 있다는 문제가 존재함
- 이 문제는 딥러닝을 이용해서 해결 할 수 있다.
 - 오타로 인해 다른 그룹으로 오판할 가능성을 줄일 수 있는 이점이 있음

S2W 데이터셋 학습 실험 결과 보고

실험 결과

- 실험 1

- F1-score (micro) 0.8934
- F1-score (macro) 0.9150
- LOC - precision: 1.0000 - recall: 0.9783 - f1-score: 0.9890
- MISC - precision: 0.8941 - recall: 0.9296 - f1-score: 0.9115
- ORG - precision: 0.8351 - recall: 0.8804 - f1-score: 0.8571
- PER - precision: 0.9688 - recall: 0.9394 - f1-score: 0.9538
- RAN - precision: 0.9613 - recall: 0.7838 - f1-score: 0.8635

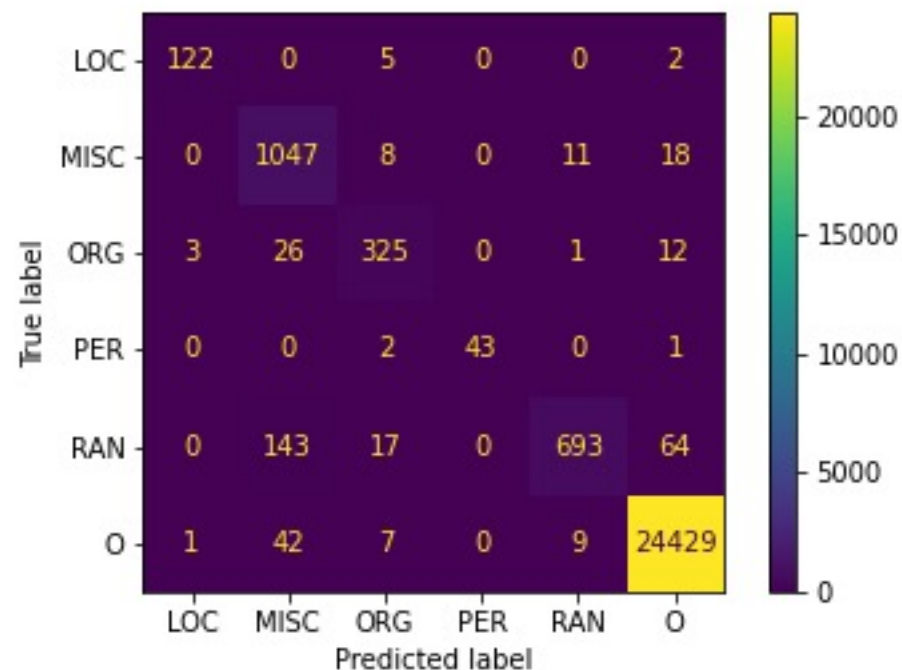


S2W 데이터셋 학습 실험 결과 보고

실험 결과

- 실험 2

- F1-score (micro) 0.8687
- F1-score (macro) 0.8946
- LOC - precision: 0.9352 - recall: 0.9619 - f1-score: 0.9484
- MISC - precision: 0.8264 - recall: 0.9656 - f1-score: 0.8906
- ORG - precision: 0.8853 - recall: 0.8655 - f1-score: 0.8753
- PER - precision: 1.0000 - recall: 0.9032 - f1-score: 0.9492
- RAN - precision: 0.9667 - recall: 0.6967 - f1-score: 0.8098

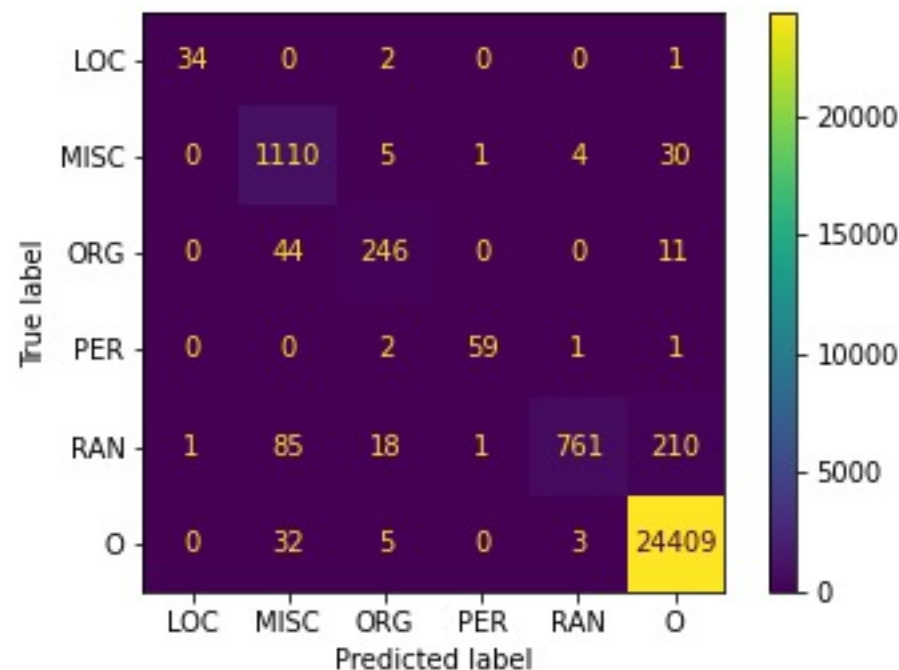


S2W 데이터셋 학습 실험 결과 보고

실험 결과

- 실험 3

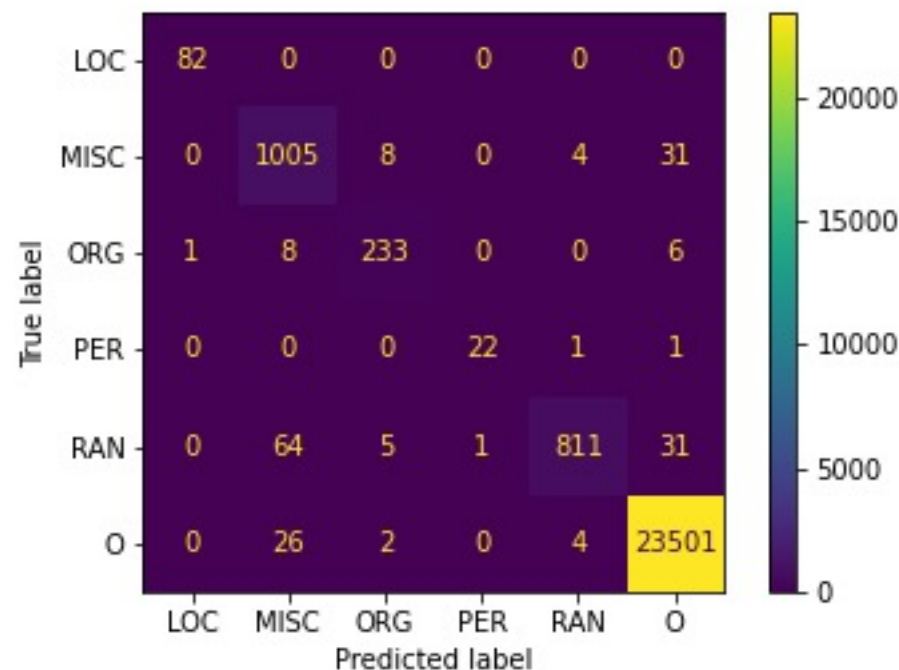
- F1-score (micro) 0.8728
- F1-score (macro) 0.8839
- LOC - precision: 0.9697 - recall: 0.9697 - f1-score: 0.9697
- MISC - precision: 0.8692 - recall: 0.9595 - f1-score: 0.9121
- ORG - precision: 0.8686 - recall: 0.7451 - f1-score: 0.8021
- PER - precision: 0.9429 - recall: 0.8919 - f1-score: 0.9167
- RAN - precision: 0.9615 - recall: 0.7132 - f1-score: 0.8190



S2W 데이터셋 학습 실험 결과 보고

실험 결과

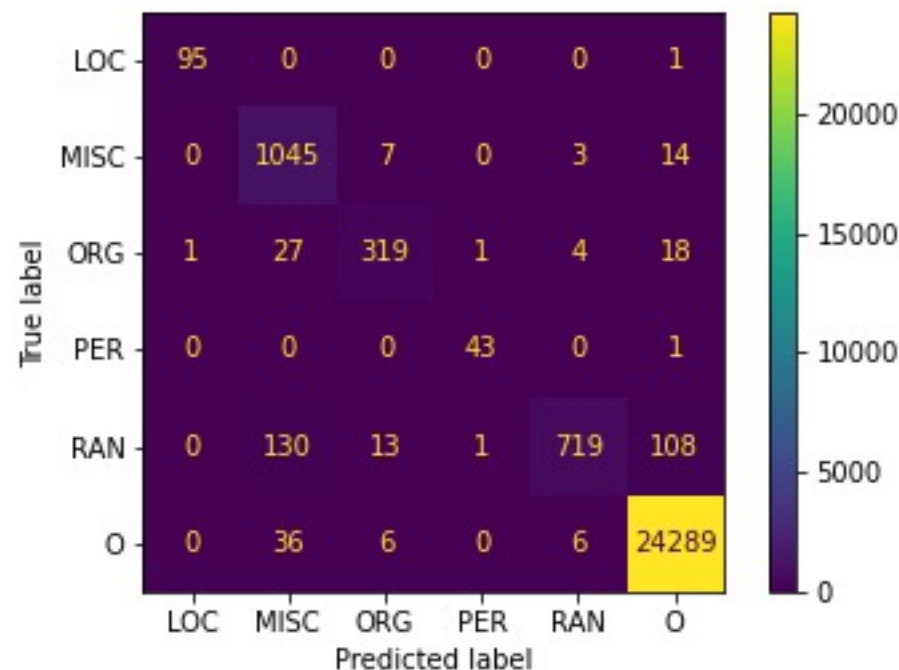
- 실험 4
 - F1-score (micro) 0.9221
 - F1-score (macro) 0.9261
 - LOC - precision: 0.9867 - recall: 1.0000 - f1-score: 0.9933
 - MISC - precision: 0.8951 - recall: 0.9665 - f1-score: 0.9294
 - ORG - precision: 0.8776 - recall: 0.8836 - f1-score: 0.8805
 - PER - precision: 0.9412 - recall: 0.8889 - f1-score: 0.9143
 - RAN - precision: 0.9844 - recall: 0.8514 - f1-score: 0.9130



S2W 데이터셋 학습 실험 결과 보고

실험 결과

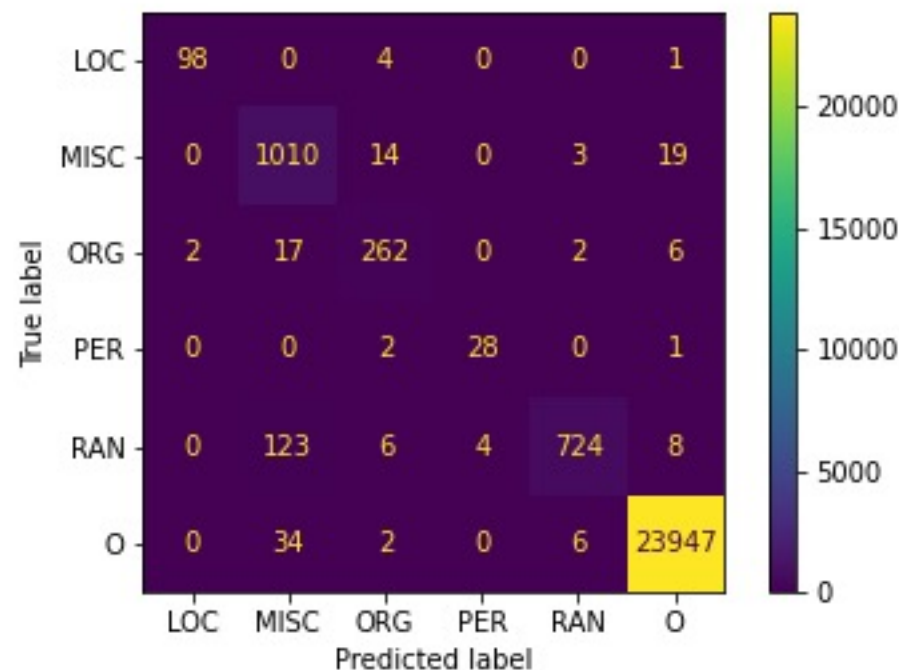
- 실험 5
 - F1-score (micro) 0.8705
 - F1-score (macro) 0.8947
- LOC - precision: 0.9390 - recall: 0.9747 - f1-score: 0.9565
- MISC - precision: 0.8409 - recall: 0.9681 - f1-score: 0.9001
- ORG - precision: 0.8841 - recall: 0.8318 - f1-score: 0.8571
- PER - precision: 0.9310 - recall: 0.9643 - f1-score: 0.9474
- RAN - precision: 0.9337 - recall: 0.7192 - f1-score: 0.8126



실험 결과

- 실험 6

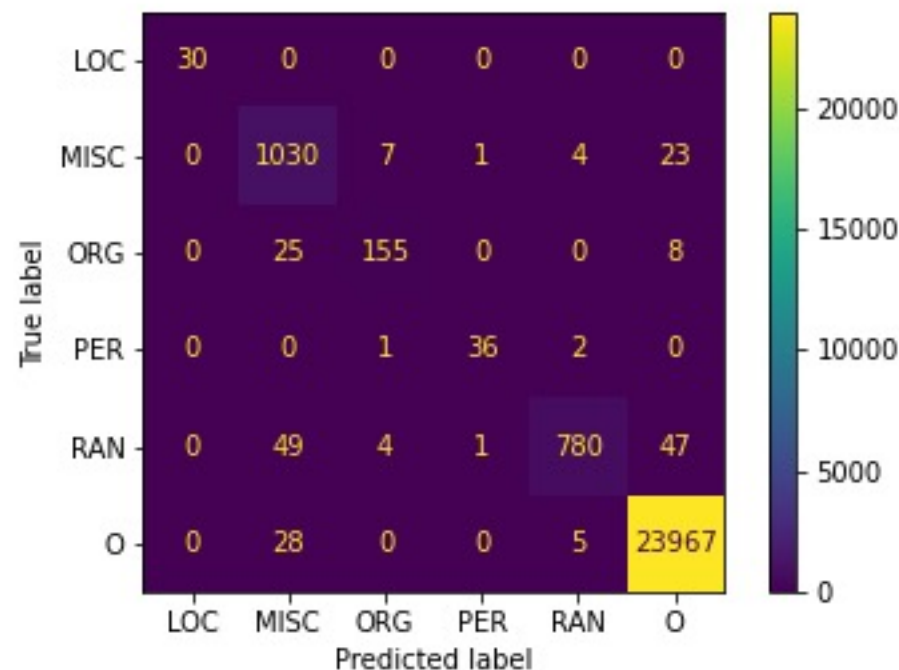
- F1-score (micro) 0.8853
- F1-score (macro) 0.8899
- LOC - precision: 0.9512 - recall: 0.9750 - f1-score: 0.9630
- MISC - precision: 0.8320 - recall: 0.9612 - f1-score: 0.8919
- ORG - precision: 0.8659 - recall: 0.8659 - f1-score: 0.8659
- PER - precision: 0.8400 - recall: 0.8750 - f1-score: 0.8571
- RAN - precision: 0.9740 - recall: 0.7883 - f1-score: 0.8714



실험 결과

- 실험 7

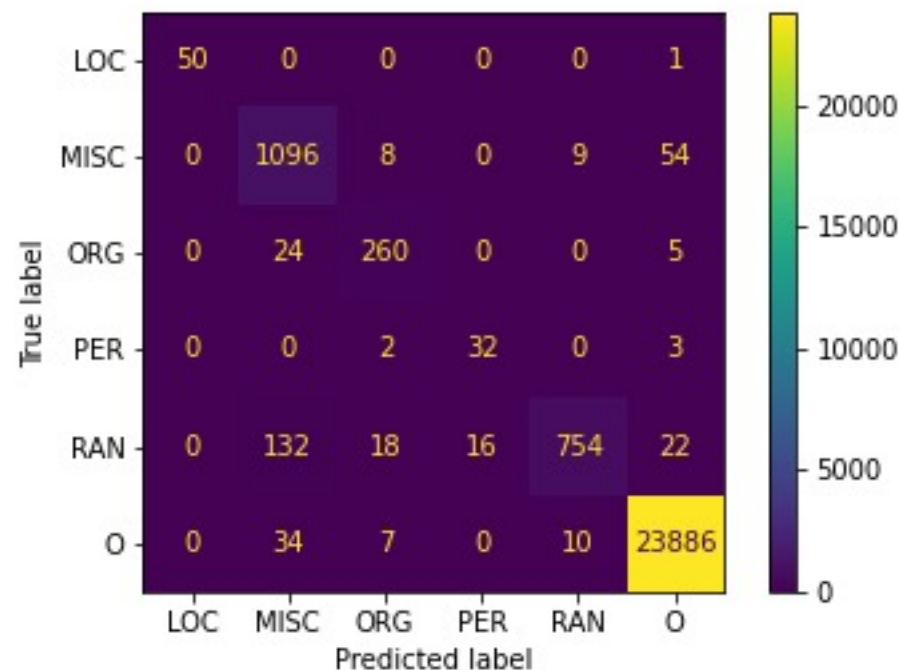
- F1-score (micro) 0.9193
- F1-score (macro) 0.9194
- LOC - precision: 1.0000 - recall: 1.0000 - f1-score: 1.0000
- MISC - precision: 0.8999 - recall: 0.9646 - f1-score: 0.9311
- ORG - precision: 0.8992 - recall: 0.7985 - f1-score: 0.8458
- PER - precision: 0.9231 - recall: 0.8889 - f1-score: 0.9057
- RAN - precision: 0.9778 - recall: 0.8589 - f1-score: 0.9145



S2W 데이터셋 학습 실험 결과 보고

실험 결과

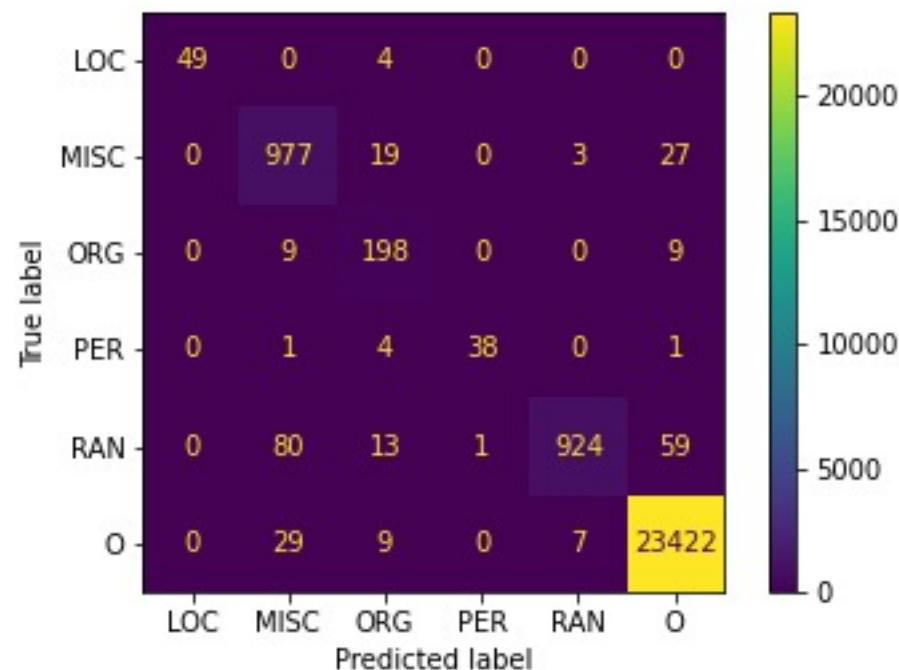
- 실험 8
 - F1-score (micro) 0.8612
 - F1-score (macro) 0.8387
- LOC - precision: 0.9512 - recall: 0.9512 - f1-score: 0.9512
- MISC - precision: 0.8385 - recall: 0.9258 - f1-score: 0.8800
- ORG - precision: 0.8413 - recall: 0.8883 - f1-score: 0.8641
- PER - precision: 0.5676 - recall: 0.8077 - f1-score: 0.6667
- RAN - precision: 0.9719 - recall: 0.7267 - f1-score: 0.8316



실험 결과

- 실험 9

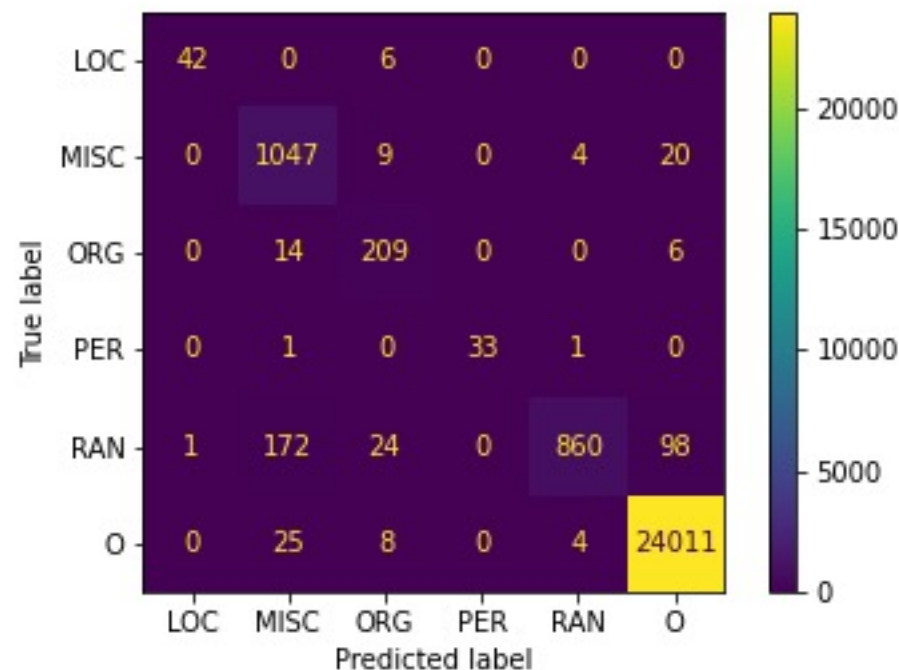
- F1-score (micro) 0.8948
- F1-score (macro) 0.8980
- LOC - precision: 1.0000 - recall: 0.9756 - f1-score: 0.9877
- MISC - precision: 0.8794 - recall: 0.9574 - f1-score: 0.9167
- ORG - precision: 0.7718 - recall: 0.8846 - f1-score: 0.8244
- PER - precision: 0.9655 - recall: 0.8235 - f1-score: 0.8889
- RAN - precision: 0.9741 - recall: 0.7898 - f1-score: 0.8723



S2W 데이터셋 학습 실험 결과 보고

실험 결과

- 실험 10
 - F1-score (micro) 0.8762
 - F1-score (macro) 0.8973
- LOC - precision: 0.9697 - recall: 0.9412 - f1-score: 0.9552
- MISC - precision: 0.8595 - recall: 0.9613 - f1-score: 0.9076
- ORG - precision: 0.8013 - recall: 0.8929 - f1-score: 0.8446
- PER - precision: 1.0000 - recall: 0.9091 - f1-score: 0.9524
- RAN - precision: 0.9338 - recall: 0.7417 - f1-score: 0.8268



S2W 데이터셋 학습 실험 결과 보고

실험 결과

F1-scores	F1-score (micro)	F1-score (macro)	LOC	MISC	ORG	PER	RAN
실험1	0.8934	0.9150	0.9890	0.9115	0.8571	0.9538	0.8635
실험2	0.8687	0.8946	0.9484	0.8906	0.8753	0.9492	0.8098
실험3	0.8728	0.8839	0.9697	0.9121	0.8021	0.9167	0.8190
실험4	0.9221	0.9261	0.9933	0.9294	0.8805	0.9143	0.9130
실험5	0.8705	0.8947	0.9565	0.9001	0.8571	0.9474	0.8126
실험6	0.8853	0.8899	0.9630	0.8919	0.8659	0.8571	0.8714
실험7	0.9193	0.9194	1.0000	0.9311	0.8458	0.9057	0.9145
실험8	0.8612	0.8387	0.9512	0.8800	0.8641	0.6667	0.8316
실험9	0.8948	0.8980	0.9877	0.9167	0.8244	0.8889	0.8723
실험10	0.8762	0.8973	0.9552	0.9076	0.8446	0.9524	0.8268
평균 ± 분산	0.8864 ± 0.0004	0.8958 ± 0.0006	0.9714 ± 0.0004	0.9071 ± 0.0003	0.8517 ± 0.0006	0.8952 ± 0.0074	0.8535 ± 0.0016

실험 결과 분석

- PER tag 분류의 분산이 크다는 문제가 있었음
 - 실험8의 경우 PER 레이블의 개별 F1스코어가 가장 나빴는데, 16개의 False Positive가 모두 랜섬웨어 이름인 Mischa를 인명으로 착각한 경우였다.
- 오류의 대부분은 RAN tag를 MISC tag로 잘못 분류한 경우였다.
 - 반대로 MISC tag를 RAN tag로 잘못 분류한 경우는 거의 존재하지 않았다.