

2. 간단한 Sentiment Analysis(감정분석)

<https://gitgutgit-ai-camp-activities-beginne-spam-ham-ui-example-oxzpwo.streamlit.app/>



2.1 스팸메일 구분, 2.2 긍정/부정 구분 (영화 리뷰 or sns 문자)

활동 내용:

- SNS 댓글 데이터나 스팸/햄 이메일 데이터셋을 사용하여 이진 분류 모델 학습.
- Scikit-learn 의 CountVectorizer, TfidfVectorizer 로 텍스트를 벡터화.
- Logistic Regression, Naive Bayes 등 간단한 모델 학습.

계획:

- 학생들이 데이터를 추가하거나 직접 레이블링 후 학습.

스팸/비 스팸 예시를 활용하여 긍정/부정 프로그램 만들기

발표 포인트

1. 이 문장이 스팸인지 아닌지 구별한다
2. “부모님이 입력한 문장이 긍정/부정인지 시가 맞춰본다.”

예시 데모 프로그램 모습:

Spam vs Ham Classifier

이 앱은 간단한 스팸/햄 분류 모델을 보여줍니다. 학생들이 부모님께 모델을 설명하고, 테스트 결과를 직접 확인할 수 있습니다.

새로운 문장 테스트하기

결과 보기

	문장	예측 결과	정답
0	Free iPhone giveaway! Hurry up, click the link!	spam	spam
1	Let's go to a movie this weekend.	spam	ham
2	Congratulations! You've been selected to win a \$1,000 gift card!	spam	spam
3	Hi jacob, see u at 4pm!	ham	ham
4	WIN a vacation to Hawaii! Limited offer, click now!	spam	spam
5	지금 구매하면 특별 할인 혜택을 드립니다! 서두르세요!	spam	spam
6	Hey, how was your day? Let's catch up soon.	ham	ham
7	오늘 날씨 좋으면 쇼핑가자	ham	ham
8	Don't forget to bring the report to the meeting tomorrow.	ham	ham
9	당신의 계정이 해킹되었습니다. 지금 복구하세요!	spam	spam

Accuracy on new texts: 90.00%

직접 문장 입력하여 테스트하기

테스트할 문장을 입력하세요:

노트북, 오늘 당장 사면 50프로 할인!

결과 확인

입력한 문장: '노트북, 오늘 당장 사면 50프로 할인!'

분류 결과: spam

3. 간단한 음성 인식 or 음성 분류 (시간이 매우 촉박하다면 시연 위주)

- 개요

- Google Speech-to-Text API 같은 것을 활용해서 음성을 텍스트로 변환 → 그 텍스트에 대해 추가 분석(감정분석 등).
- 또는 음성의 특성(스펙트로그램)으로 간단한 분류(예: 음악 장르 분류)를 시연.

주의점

- 음성 인식은 외부 API 사용하면 간단하지만, 직접 딥러닝 모델을 학습하려면 시간이 많이 걸림(데이터셋, GPU 등).

- 발표 포인트

- “이게 머신러닝(딥러닝)이 실제 음성을 어떻게 처리해 텍스트 또는 카테고리를 추출하는지” 간단한 원리 소개