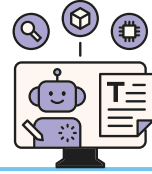


# 03 텍스트 데이터를 이용한 영화 평가 시스템

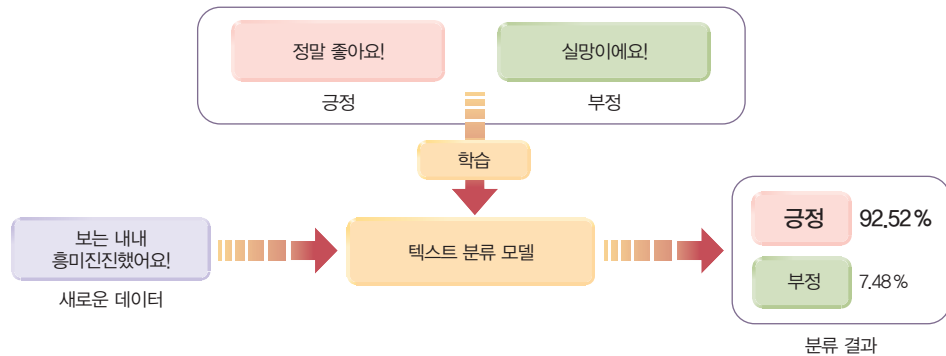
## 학습 목표

- 텍스트 데이터를 분류하는 인공지능 모델을 생성할 수 있다.
- 텍스트 분류 모델을 활용하여 영화 평가 시스템을 구현할 수 있다.



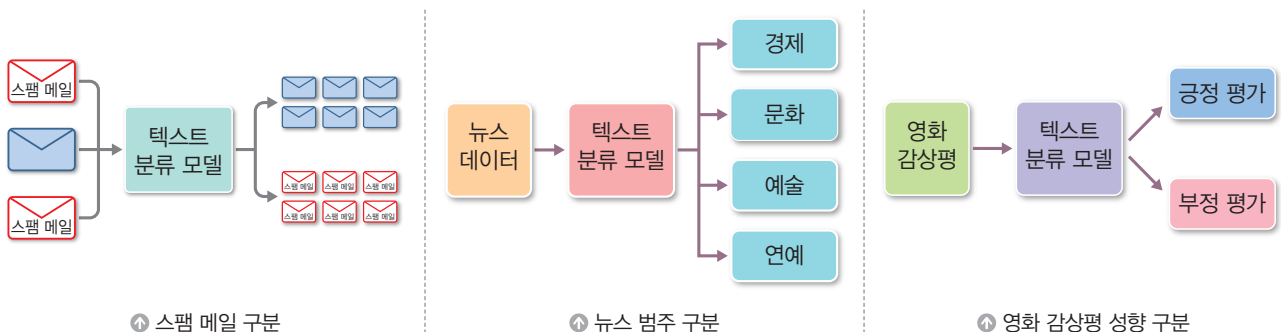
## 1 텍스트 데이터를 분류하는 인공지능

텍스트 분류 모델은 텍스트 데이터를 학습하는 과정에서 텍스트의 패턴을 분석하여 특정 텍스트가 어떤 범주나 클래스에 해당하는지 분류하는 모델이다. 예를 들어 영화 댓글 평가 시스템에 긍정과 부정의 댓글 데이터를 충분히 학습시킨 텍스트 분류 모델을 이용하면 댓글의 전체적인 반응을 알 수 있다.



[그림 IV-12] 텍스트 데이터의 분류 예시

텍스트 분류 모델은 이메일 필터링 시스템에서 일반 메일과 스팸 메일을 구분하거나 사회 관계망 서비스(SNS)에서 사용자의 글을 분석하여 부적절한 글을 찾아내는 데에도 활용할 수 있다.



[그림 IV-13] 텍스트 분류 모델의 활용

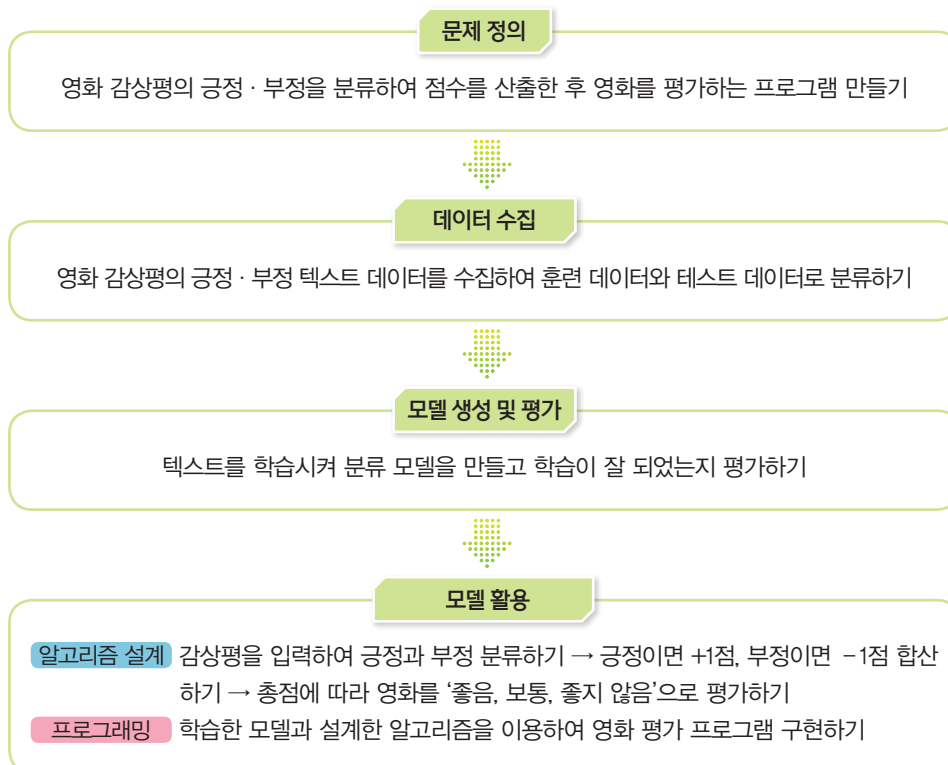
## 2 문제 파악하기

같은 영화를 보더라도 사람에 따라 감상평이 서로 달라 전반적인 평가를 파악하기 어려울 때가 있다. 영화의 감상평을 긍정과 부정으로 분류하여 전반적인 반응이 어떠한지 알려 주는 영화 평가 시스템을 만들어 보자.



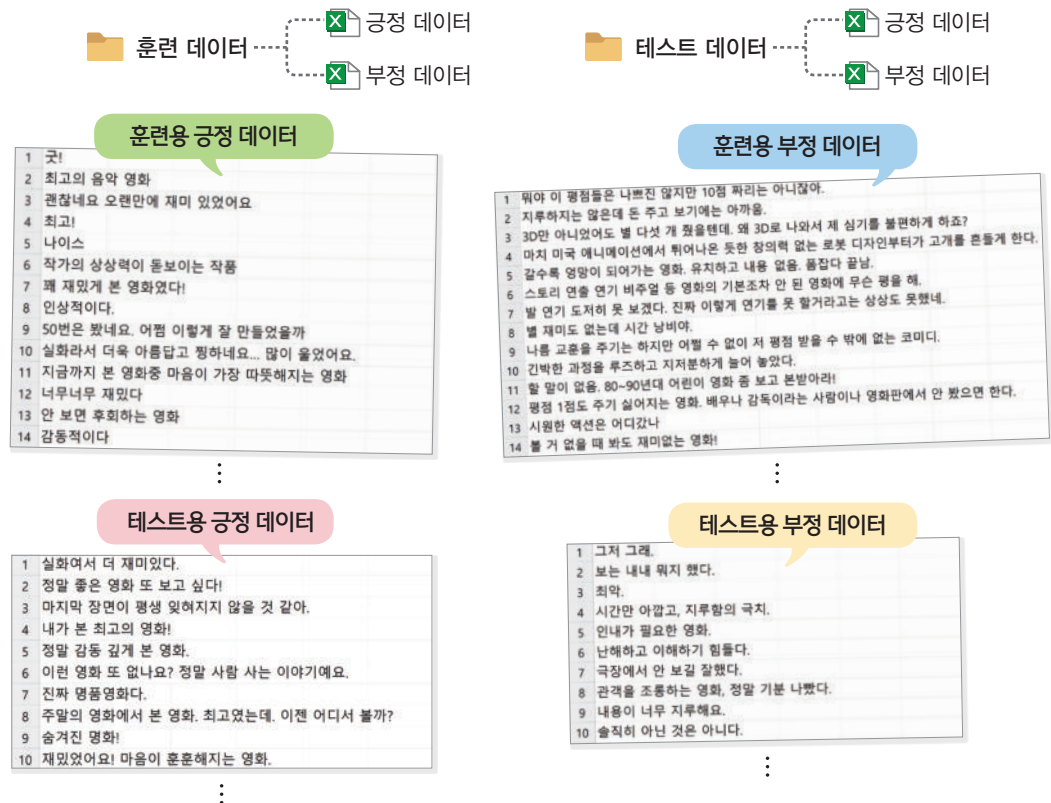
### 문제 해결 과정

텍스트 분류 모델을 활용한 영화 평가 시스템을 만드는 과정은 다음과 같다.



### 3 데이터 수집

인공지능 모델에 감상평을 학습시키기 위해 긍정적인 평가와 부정적인 평가에 해당하는 글을 수집하고 적절하지 않은 데이터는 삭제한 후 레이블링한다. 수집한 감상평을 다음과 같이 학습에 사용할 훈련 데이터와 학습이 잘 되었는지 확인할 때 사용할 테스트 데이터로 분류하여 저장한다.



### 4 모델 생성 및 평가

준비한 데이터를 학습시켜 감상평이 긍정인지 부정인지 구분하는 분류 모델을 생성한 후 평가한다.

① 엔트리의 **인공지능** 블록 꾸러미에서 **인공지능 모델 학습하기**를 클릭한 후, [새로 만들기] - [분류: 텍스트] - **학습하기** 메뉴를 차례대로 선택한다.



② 모델명과 클래스명을 다음과 같이 설정한다.

모델명

클래스명

**분류: 텍스트 모델 학습하기**

**영화 평가**

**데이터 입력**  
모델이 학습할 데이터를 입력합니다.

긍정 0개 X

모델이 학습할 텍스트 데이터를 아래에 작성해 주세요.  
클래스 당 5개 이상의 데이터를 입력해야 합니다.  
각각의 데이터는 심표로 구분합니다. (예: 맛있다, 맛있어, 맛있네)

클릭해서 데이터를 입력해 주세요.

부정 0개 X

모델이 학습할 텍스트 데이터를 아래에 작성해 주세요.  
클래스 당 5개 이상의 데이터를 입력해야 합니다.  
각각의 데이터는 심표로 구분합니다. (예: 맛있다, 맛있어, 맛있네)

클릭해서 데이터를 입력해 주세요.

③ 클래스마다 준비한 데이터 파일을 업로드한다.

긍정 0개 X

모델이 학습할 텍스트 데이터를 아래에 작성해 주세요.  
클래스 당 5개 이상의 데이터를 입력해야 합니다.  
각각의 데이터는 심표로 구분합니다. (예: 맛있다, 맛있어, 맛있네)

클릭해서 데이터를 입력해 주세요.

➡

긍정 0개 X

모델이 학습할 텍스트 데이터를 아래에 작성해 주세요.  
클래스 당 5개 이상의 데이터를 입력해야 합니다.  
각각의 데이터는 심표로 구분합니다. (예: 맛있다, 맛있어, 맛있네)

텍스트 데이터를 작성해 주세요.

파일 업로드  
1GB 이하의 txt, csv 형식의 파일을 업로드할 수 있습니다.  
업로드할 파일의 내용은 데이터의 첫 마다에 추가됩니다.

➡

긍정 56개 X

모델이 학습할 텍스트 데이터를 아래에 작성해 주세요.  
클래스 당 5개 이상의 데이터를 입력해야 합니다.  
각각의 데이터는 심표로 구분합니다. (예: 맛있다, 맛있어, 맛있네)

곳!, 최고의 음악 영화, 괜찮네요 오랜만에 재미 있었어요, 최고!, 나이스, 작가의 상상력이 돋보이는 작품, 꽤 재미있게 본 영화였다!, 인상적이다., 50번은 봤네요, 어쩔 어쩔하게 볼 만할것을嘛, 실화라서 더욱 아름답고 정하네요~ 말이 많았어요., 지금까지 본 영화중 마음에 가장 따뜻한 영화, 너무너무 재밌다. 안 보면 후회하는 영화, 감동적이다., 아른스타일 액션영화 굿굿!!, 아이다이어가 아

파일 업로드  
1GB 이하의 txt, csv 형식의 파일을 업로드할 수 있습니다.  
업로드할 파일의 내용은 데이터의 첫 마다에 추가됩니다.

텍스트 데이터는 심표(.)로 구분한다.

④ [모델 학습하기] 버튼을 클릭하여 학습을 진행한다.

**분류: 텍스트 모델 학습하기**

모델 학습은 인터넷이 연결되어 있어야 정상적으로 동작합니다. [문제리포트 보기](#)

**영화 평가**

**데이터 입력**  
모델이 학습할 데이터를 입력합니다.

긍정 61개 X

모델이 학습할 텍스트 데이터를 아래에 작성해 주세요.  
클래스 당 5개 이상의 데이터를 입력해야 합니다.  
각각의 데이터는 심표로 구분합니다. (예: 맛있다, 맛있어, 맛있네)

곳!, 최고의 음악 영화, 어쩔의 작품위에 참여오는 새로운 인연의 기쁨, 괜찮네요, 오랜만에 재미 있었어요, '눈에 보이는' 반영이었지~

부정 61개 X

모델이 학습할 텍스트 데이터를 아래에 작성해 주세요.  
클래스 당 5개 이상의 데이터를 입력해야 합니다.  
각각의 데이터는 심표로 구분합니다. (예: 맛있다, 맛있어, 맛있네)

뭐야 이 형들은 나쁜건 많지만 10점까지는 아니잖아, 지루하지는 않은데 돈 주고 보게는 아까워., 30만 아니었어도 몇 다섯 개 할~

**학습**  
입력한 데이터로 모델을 학습합니다.

모델 학습하기

모델 학습을 완료했습니다. 100%

학습이 완료되면 메시지 표가 표시된다.

**결과**  
학습한 모델의 결과를 확인합니다.

입력하기

분류한 클래스  
긍정

테스트 데이터를 업로드하여 학습 결과 확인

**적용하기** → 학습된 모델 생성

- 5 모델이 충분히 학습하였는지 확인하기 위하여 [결과]의 입력 창에 테스트 감상평을 입력하여 결과를 확인한다. 분류가 잘 되었다면 [입력하기]를 클릭하여 프로그래밍에 활용한다.

결과  
학습한 모델의 결과를 확인합니다.

이런 영화 또 없나요? 정말 사람 사

입력하기

분류한 클래스  
**긍정**

결과  
학습한 모델의 결과를 확인합니다.

보는 내내 뒹지 했다.

입력하기

분류한 클래스  
**부정**

▲ 테스트 데이터로 텍스트 분류 모델 학습 결과 확인하기

## 5 모델 활용

텍스트 분류 모델을 활용하여 영화 평가 시스템을 만들기 위한 알고리즘을 설계하고 이를 이용하여 프로그래밍한다.

### 알고리즘 설계하기



- 1 입력받은 값이 1이 될 때까지 영화 감상평을 입력받아 긍정인지 부정인지 분류한다.
- 2 분류 결과가 긍정이면 '점수' 변수에 1을 더하고, 부정이면 '점수' 변수에 -1을 더한다.
- 3 평균에 따라 감상평이 좋은 영화, 보통인 영화, 좋지 않은 영화로 알려 준다.

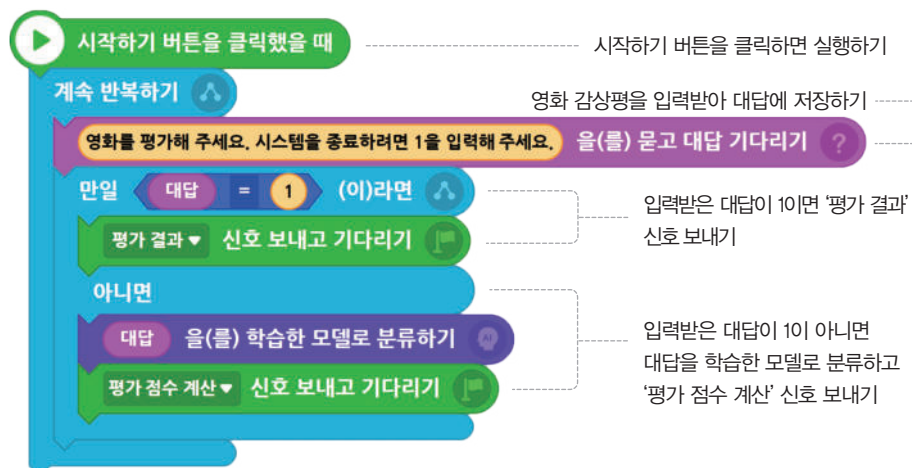
### 프로그래밍하기



- 1 입력받은 값이 1이 될 때까지 영화 감상평을 입력받아 긍정인지 부정인지 분류한다.



엔트리봇





2 분류 결과가 긍정이면 '점수' 변수에 1을 더하고, 부정이면 '점수' 변수에 -1을 더한다.



점수를 영화 감상평 입력 개수로 나누는 이유

점수를 영화 감상평 입력 개수로 나누는 이유는 점수 범위를 -1.0 ~ 1.0으로 분포시킨 후 0.5 ~ 1.0은 평가가 좋은 영화, -0.5 ~ 0.5는 평가가 보통인 영화, -1.0 ~ -0.5는 평가가 좋지 않은 영화로 구분하기 위해서이다.

3 평균에 따라 감상평이 좋은 영화, 보통인 영화, 좋지 않은 영화로 알려 준다.



엔트리봇

## 1 영화 감상평 분류 프로그램의 예측 결과를 확인해 보자.

영화평	실제 정답	예측 결과
내가 본 영화 중 최고예요!	긍정	
재미있어요. 마음이 훈훈해져요.	긍정	
한 번 더 볼 거예요.	긍정	
지루함의 극치입니다.	부정	
돈이 아까운 영화입니다.	부정	
기대했지만 별로예요.	부정	

## 2 데이터의 영화평을 읽어 보고 1~5의 점수로 구분한 후, 인공지능 모델을 학습시켜 보자.

(1: 매우 부정, 2: 부정, 3: 보통, 4: 긍정, 5: 매우 긍정)

(1) 모델을 수정해 보자.

(2) 프로그램을 수정해 보자.

## 3 수정한 프로그램이 다음의 영화평에 몇 점의 평점을 주는지 확인해 보자.

영화평	실제 평점	예측 평점
영화의 반전이 대단했다. 실망시키지 않는 감독이다.	5	
연기가 아쉽지만 재미있었다.	4	
이전 작품보다 별로였다. 조금 뻘한 스토리다.	2	
오랜만에 재미있는 공포 영화였습니다.	5	

다음 항목을 점검하여 인공지능 윤리를 지키고 있는지 나와 친구의 프로그램을 확인해 보자.

평가 항목	자기 평가	동료 평가
인공지능을 활용하려는 목적이 올바른가?	예 <input type="checkbox"/> 아니오 <input type="checkbox"/>	예 <input type="checkbox"/> 아니오 <input type="checkbox"/>
자료를 수집할 때 개인 정보나 저작권을 보호하도록 하였는가?	예 <input type="checkbox"/> 아니오 <input type="checkbox"/>	예 <input type="checkbox"/> 아니오 <input type="checkbox"/>
인공지능 모델을 학습시킬 때 편향된 결과를 산출할 가능성을 최소화하였는가?	예 <input type="checkbox"/> 아니오 <input type="checkbox"/>	예 <input type="checkbox"/> 아니오 <input type="checkbox"/>
인공지능 모델을 활용하는 프로그램이 다른 사람에게 피해를 줄 가능성은 없는가?	예 <input type="checkbox"/> 아니오 <input type="checkbox"/>	예 <input type="checkbox"/> 아니오 <input type="checkbox"/>