

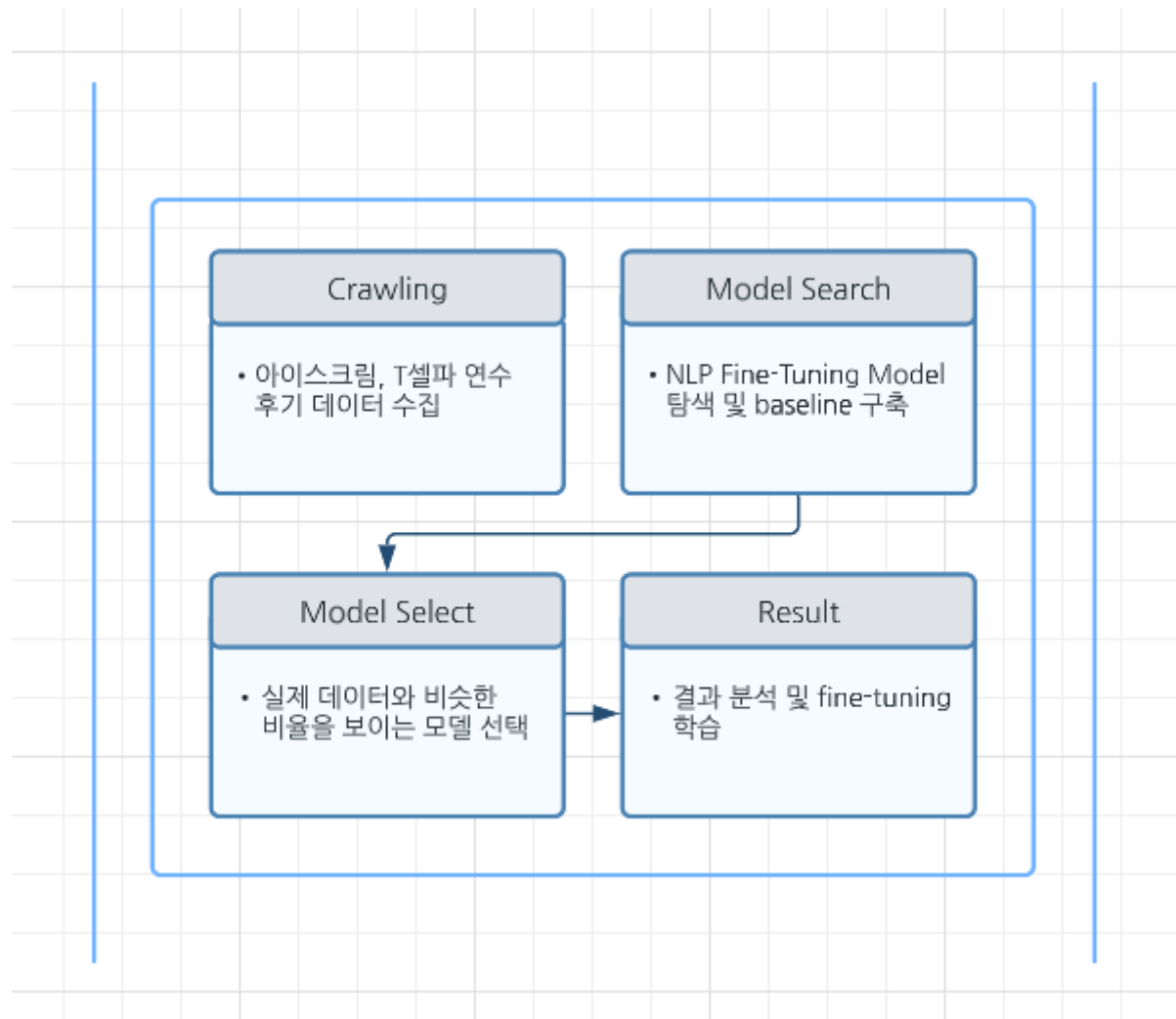
프로젝트 분석 보고서

: T 셀파, i-Scream 연수후기 비교분석

프로젝트 기반 빅데이터 서비스 개발자 양성 과정 (4 기)

이형석

개요



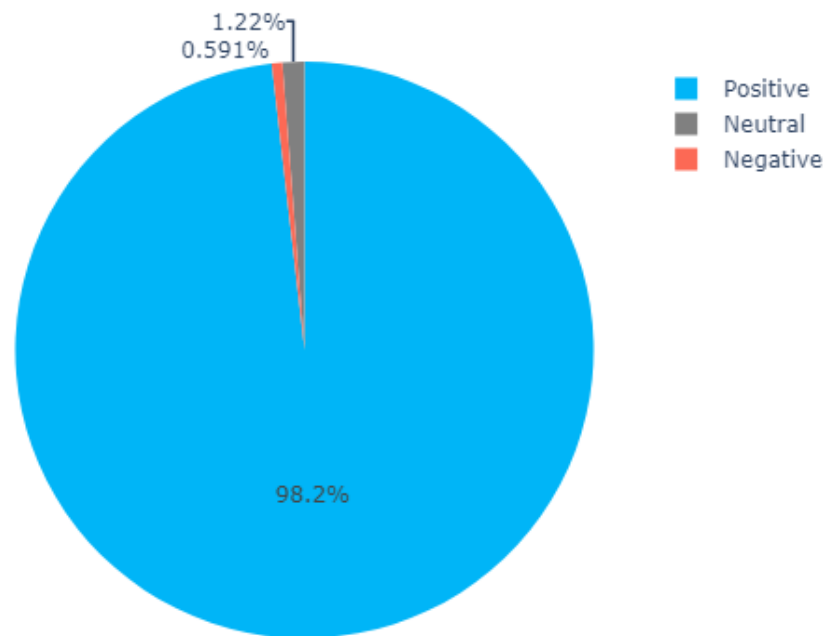
실험

- 모델 탐색

후기를 작성한 사람들이 선택한 별점과 Pre-Trained 모델이 분류하는 비율은 어떻게 다를지 실험했다. 별점은 3 점을 중립 점수로 설정하고 별점 4,5 개는 긍정, 별점 1,2 개는 부정으로 나누었다. 위 그림을 보면 수집한 데이터에서 98.2%의 높은 비율로 긍정의 평가가 많다는 것을 알 수 있다.

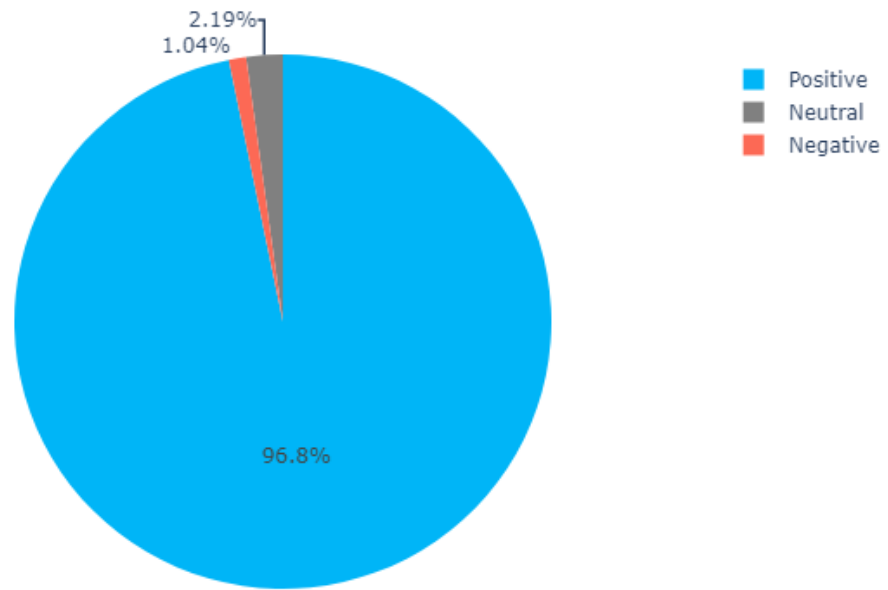
Huggingface 를 참고하여 제작자와 모델명을 적음으로써(ex. klue/bert-base) 비교적 간단히 Pre-Trained 모델을 불러왔다. 그 다음 AutoTokenizer 와 AutoModelForSequenceClassification, TextClassificationPipeline 으로 수집한 데이터들의 감성 분류를 진행했다.

- 시각화



[i-Scream] Sentimental Classification by Star (%)

아이스크림 Positive, Neutral, Negative 감정 분석 비율



[T-sherpa] Sentimental Classification by Star (%)

T 셀파 Positive, Neutral, Negative 감정 분석 비율

결과

[‘matthewburke/korean_sentiment’](#) 모델

아이스크림		T셀파	
Positive	5,108 (97.5%)	Positive	2,526 (97.2%)
Neutral	68 (1.3%)	Neutral	25 (1.88%)
Negative	65 (1.24%)	Negative	49 (0.962%)

klue/bert-base 모델

아이스크림		T셀 파	
Positive	5230	Positive	0
Neutral	9	Neutral	2600
Negative	2	Negative	0

skt/kogpt2-base-v2 모델

아이스크림		T셀 파	
Positive	0	Positive	0
Neutral	5,182	Neutral	2,569
Negative	59	Negative	31

beomi/kcbert-large 모델

아이스크림		T셀 파	
Positive	5,064	Positive	0
Neutral	172	Neutral	2,549
Negative	5	Negative	51

작업 코드

https://github.com/lhshs/mini_pjt/tree/main/230727-01_NLP