

2. Paperswithcode(<https://paperswithcode.com/area/natural-language-processing>)에서 NLP sub task 중에 2개를 선택하여 본인 블로그에 정리해보세요. task 별로 아래 3가지 항목에 대해서 정리하세요. (각 항목 고려 사항 참고)

- 문제 정의
  - task가 해결하고자 하는 문제가 무엇인가?

- 데이터 소개(대표적인 데이터 1개)

- task를 해결하기 위해 사용할 수 있는데 데이터가 무엇인가?
  - 데이터 구조는 어떻게 생겼는가?

- SOTA 모델 소개(대표적인 모델 1개)

- task의 SOTA 모델은 무엇인가?
  - 해당 모델 논문의 요약에서 주요 키워드는 무엇인가?

- <https://blog.naver.com/ctk456/222653835941>

POB

## Sub task of NLP(Sentiment Analysis / Text Generation)

아기벨 · 1시간 전

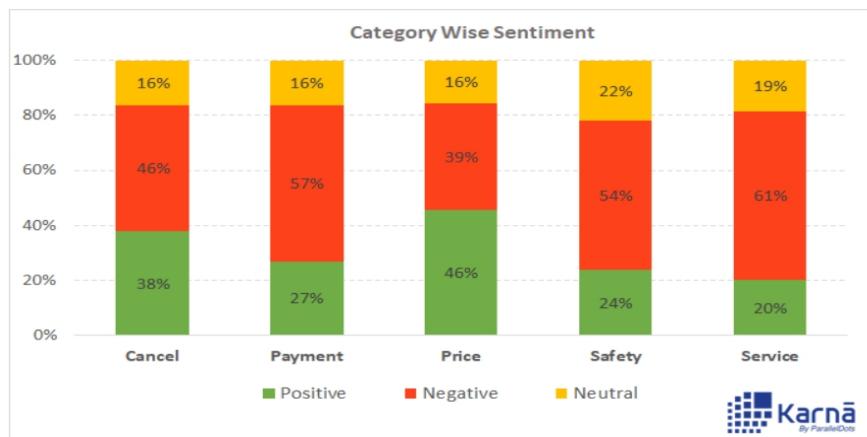
URL 복사

번역

### Sentiment Analysis(감성분석)

- 감성분석은 자료(주로 텍스트)에서 주관적인 정보를 파악해 긍정, 부정, 중립 등 감정을 추출해내는 분석으로 오피니언 마이닝(Opinion Mining)이라고 불리기도 한다.

브랜드의 제품이나 서비스의 만족도를 파악하기 위해 사용된다.



예) 우버의 카테고리에 대한 감성분석 ([클릭](#))

### 감성분석에서 사용하는 데이터

- 대표적인 데이터셋으로 ML에 titanic이 있다면 감성분석에는 IMDb 영화리뷰가 있다.

**IMDb(Internet Movie Database)** : 긍정 또는 부정으로 라벨링 된 5만여 개의 이진분류 데이터 셋  
IMDb의 특징은 데이터의 구조는 아래 페이지에서 확인할 수 있다.

#### IMDb Dataset Details

Each dataset is contained in a gzipped, tab-separated-values (TSV) formatted file in the UTF-8 character set. The first line in each file contains headers that describe what is in each column. A 'NaN' is used to denote that a particular field is missing or null for that title/name. The available datasets are as follows: