

Part. 01

실전 데이터 분석 프로젝트

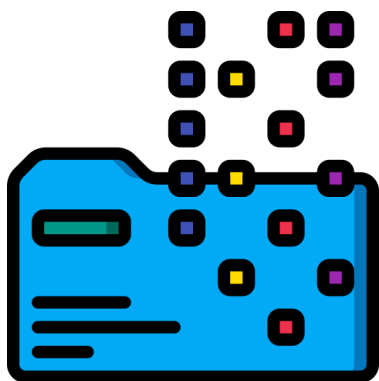
| 텍스트마이닝을 활용한 카카오톡 토픽분석

FASTCAMPUS
ONLINE

머신러닝과 데이터분석A-Z

강사. 김용담

I 텍스트 마이닝 과정



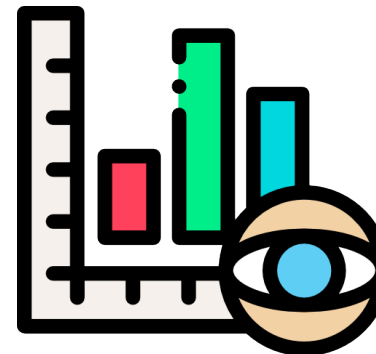
텍스트 수집



텍스트 전처리

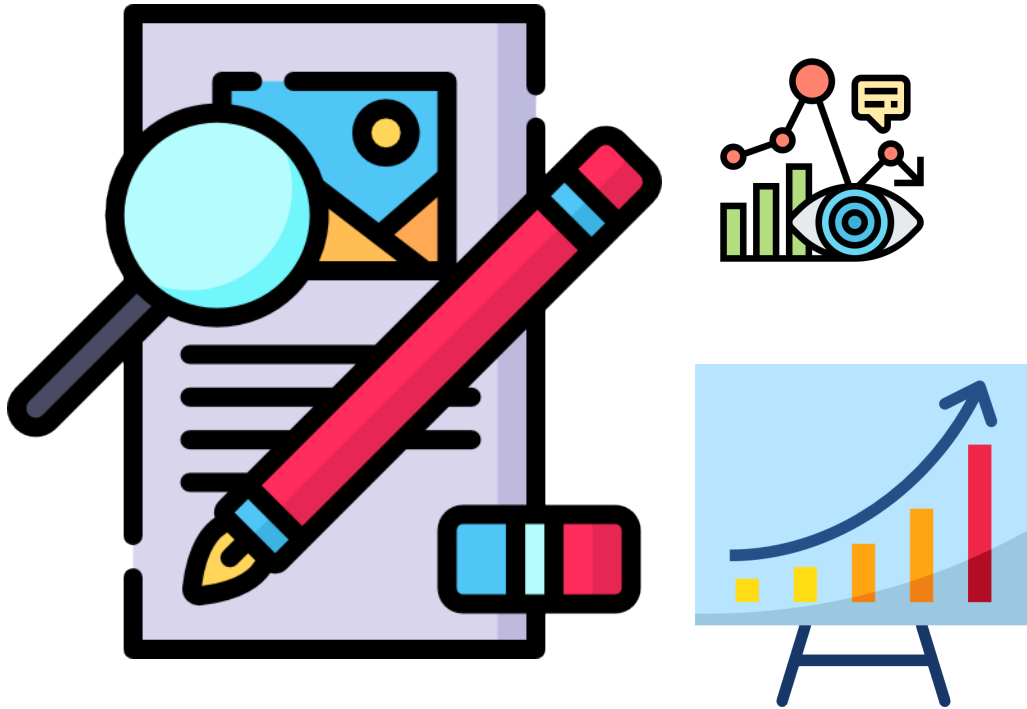


텍스트 분석



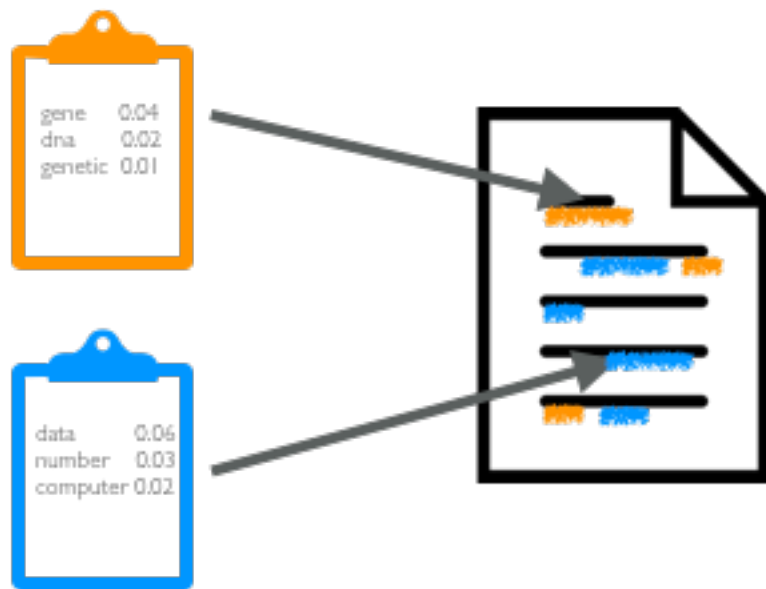
텍스트 시각화

I 텍스트 분석



- 텍스트가 가지고 있는 숨어있는 의미를 파악하는 모든 기법
- 언어 모델, 머신 러닝 기법, 통계적 분석 기법들이 쓰임
- 정보 검색, 토픽 모델링, 감정 분석, 문서 분류 등이 있음

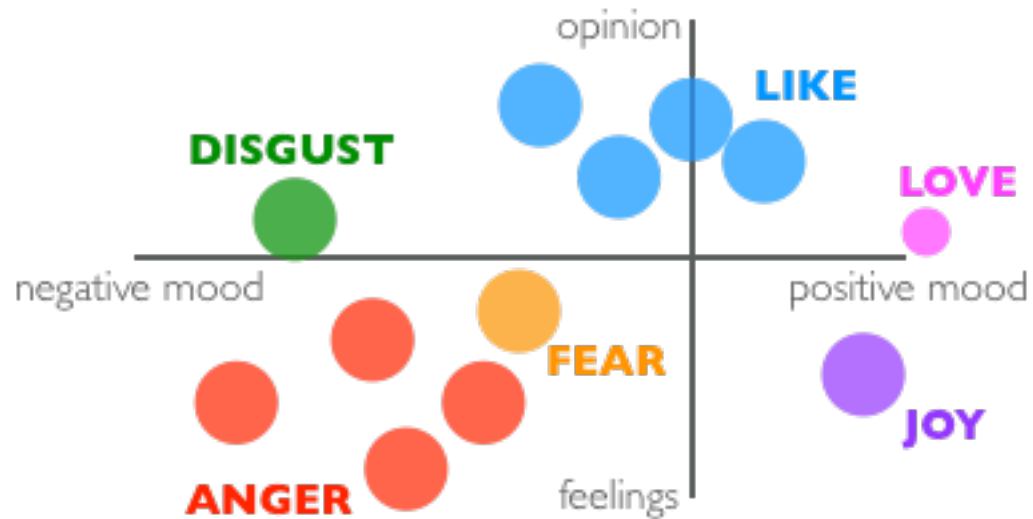
I 대표 예시 1. 토픽 모델링(Topic Modeling)



토픽 모델링

- 문서들이 가지고 있는 여러가지 토픽을 찾아주는 기법
- 토픽을 단어의 묶음으로 정의하고, 하나의 문서에 여러 개의 토픽이 있을 수 있다고 가정
- 토픽 모델링의 결과로 모든 단어들이 각 토픽에 속할 확률이 계산되어 나옴
- 대표적인 방법으로, Latent Dirichlet Allocation(LDA), Author Topic Model(ATM), Dynamic Topic Model(DTM) 등이 있음

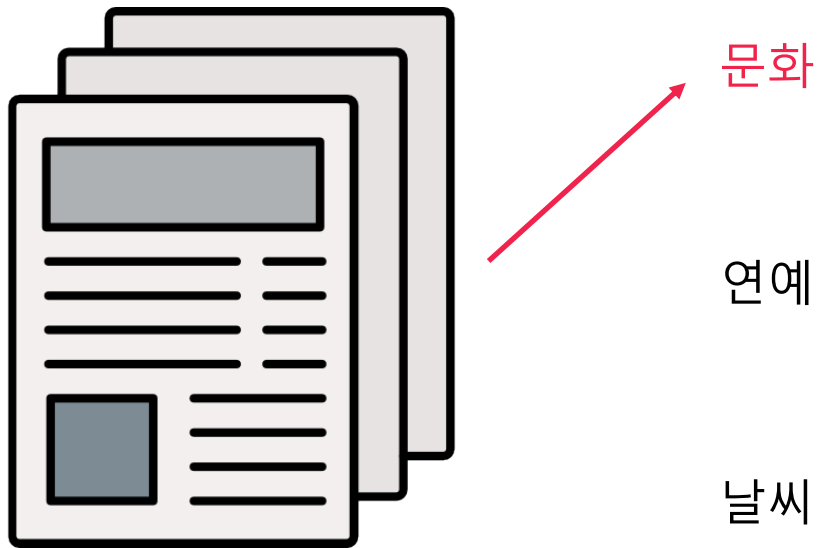
I 대표 예시 2. 감정 분석(Sentiment Analysis)



드라마 의견 분석

- 텍스트에 포함된 저자의 의견을 찾아내는 기법
- 가장 쉬운 예시로는 영화 리뷰 댓글의 '긍정' 또는 '부정'을 파악하여 평가하는 모델이 있음
- 확정된 개념으로 "Opinion Mining" 이라고도 불림
- 지도 학습의 개념으로 텍스트 분류 방법으로 사용되거나, 비지도 학습의 개념으로 문맥을 파악하여 의견을 찾아내는 방법이 있음

I 대표 예시 3. 텍스트 분류(Text Classification)



- 텍스트를 지정한 카테고리로 분류하는 기법
- 텍스트 분석 방법 중 가장 많이 사용되는 기법이며, 대부분 지도 학습 방법을 사용
- 과거에는 머신 러닝 기법들을 많이 사용
예) Support Vector Machine, Random Forest, Naïve Bayes Classifier 등
- 현재는 딥러닝 모델을 많이 사용
예) CNN, RNN, LSTM, multi-layer RNN, Bi-LSTM 등

Part. 01

실전 데이터 분석 프로젝트

I 감사합니다.

FASTCAMPUS
ONLINE

머신러닝과 데이터분석A-ZI

강사. 김용담