

# 텍스트 마이닝 프로세스

텍스트 마이닝은 데이터 수집, 전처리, 인사이트 도출 과정입니다. 이 프로젝트는 데이터 전처리, 단어 벡터 생성, 토픽 모델링, 감성 분석을 순차적으로 수행했습니다.



작성자: sanggoo cho



# 텍스트 마이닝 전체 프로세스



1

## 데이터 분석

텍스트 데이터 정제, 결측치 처리, 탐색적 데이터 분석 수행

2

## 단어 벡터 생성

단어 추출, 벡터화하여 머신러닝 모델에서 사용 가능한 형태로 변환

3

## 토픽 모델링

벡터화된 데이터로 잠재된 주제 분석, 텍스트 데이터의 주제 분포 파악

4

## 감성 분석

텍스트 데이터를 긍정, 부정, 중립으로 분류하여 감성적 경향 파악



# 데이터 분석

## 1 데이터 불러오기 및 결측치 처리

데이터셋 로드 후 결측값 제거

## 2 텍스트 정제

특수 문자와 불필요한 공백 제거

## 3 탐색적 데이터 분석

데이터 분포와 특성별 통계적 특징 시각화

# 단어 벡터 생성

## 단어 추출

Okt 라이브러리로 명사 추출, 불용어 제거하여 코퍼스 구성

## TF-IDF 생성

CountVectorizer와 TfidfVectorizer로 단어 벡터와 TF-IDF 매트릭스 생성

## 워드클라우드 생성

단어 벡터로 워드클라우드 시각화, 주요 키워드 표현





# 토픽 모델링 - LDA 분석

## LDA 적용

LatentDirichletAllocation으로 코퍼스  
퍼스를 주제로 분류

## 주제 추출

텍스트 데이터에서 잠재된 주제 추출

## 결과 해석

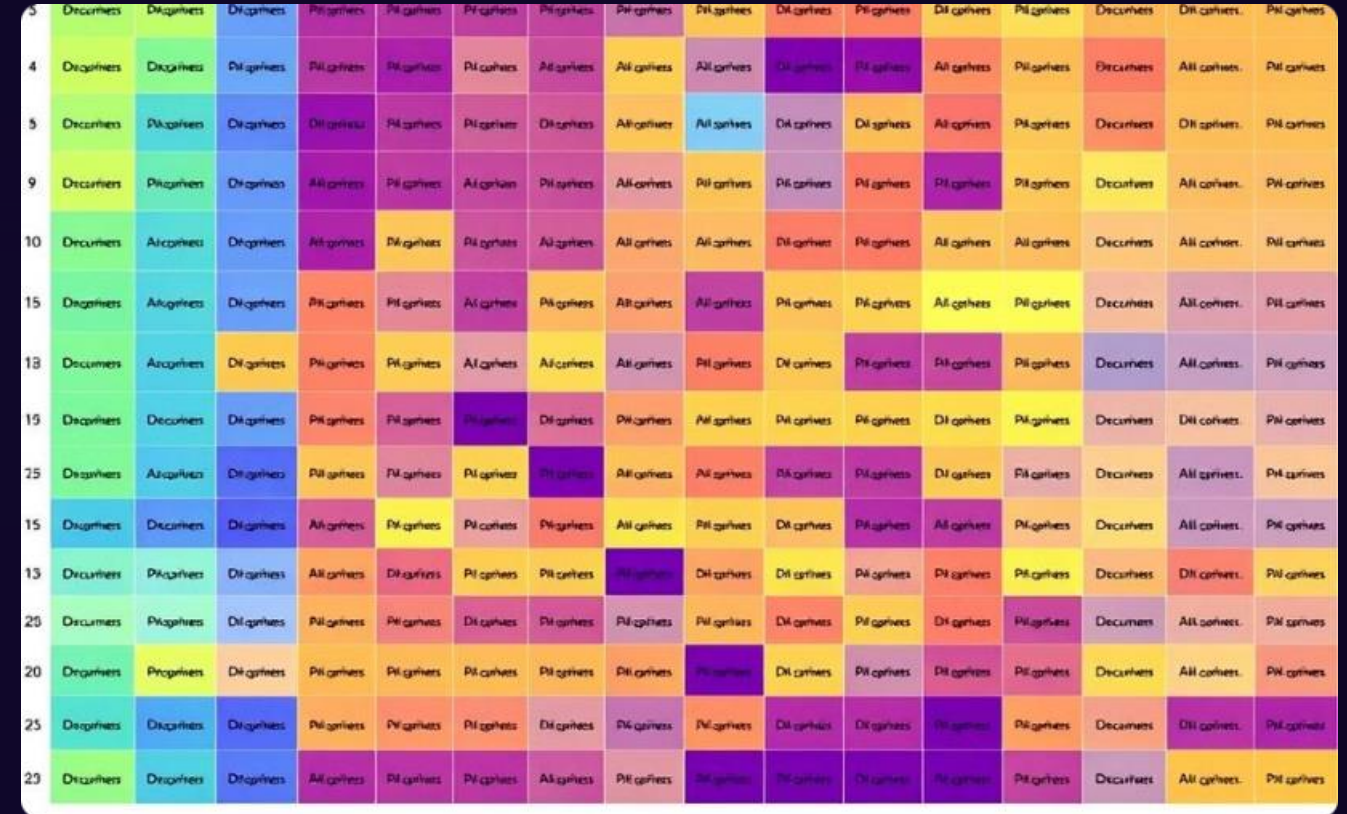
각 문서가 어느 토픽에 속하는지 확인

# 토픽 모델링 - 시각화



# pyLDAvis 시각화

## 각 주제를 대표하는 단어들을 시각적으로 분석



## 토픽 분포 히트맵

문서별 토픽 분포를 색상으로 표현한 히트맵

# 감성 분석 - 모델 적용

1

## 모델 로드

Transformers 라이브러리의 pipeline 함수로 감성 분석 모델 로드

2

## 감성 점수 계산

특정 문장에 대한 감성 점수 계산

3

## 감성 분류

문장을 긍정, 부정, 중립으로 분류





# 감성 분석 - 결과 처리



## 결과 저장

감성 분석 결과를 리스트에 저장



## 결과 출력

각 문장의 감성과 점수 출력



## 결과 분석

출력된 결과를 바탕으로 데이터의 감성 경향 분석

ntcattstin Rarlyst



\$1698

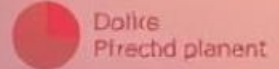
135 roes 3tn  
Caper ariuk

Sentiment analysiess sentition gharts



34 1298

Trp2



6+1%

164,26

