

4차시 가이드라인

팀 구성/ 팀장 정하기/ 데이터, 도메인지식 조사

생각해 보기

자연어가 무엇이고 어디에서 활발하게 사용되고 있는지?
이러한 이슈들을 봤을때 프로젝트가 어떤 방향으로 가면 좋을지?

칼럼 확인해 보기

칼럼 살펴보면 **string** 이런식으로 되어있는데 이게 무엇을 뜻하는지?

왜 각 칼럼의 값이 이렇게 적혀있는지 생각해보기

과정 이해해 보기

데이터의 종류 (정형/비정형데이터)란?

텍스트 마이닝이란?

Corpus(말뭉치)란?

Token 란?

Stemming 란?

n-gram 란?

Stopword(불용어) 란?

텍스트 마이닝 과정은?

텍스트 마이닝 전처리란?