<중간발표 피드백>

기업개요 데이터에 대한 회의

Ex) 삼성전자 같은 경우 너무 많은 분야가 있다.

마지막 발표 때는 구체적인 투자 계획을 말해 줄 것.

클러스터 내의 기업간 유사도에 집중해서 전략 세워보자.

------------------------------------------------------------------------------------------------------------

동헌

벡터화 방법

One hot encoding

Tfidf count

\*Word2vec 단어의 맵핑, / skip gram 주변 단어로 1개의 중심 단어

Voca -> countvectorizer 에서 뽑은 단어로 하자

우리가 지정해주자

은수 -<< 사전 과 stop\_words\_ 공유해주기

형태소 사전

재무지표 데이터 (DL 이앤씨, LX 홀딩스는 제외 아예 데이터가 없음)

여러 알고리즘 실험 해보기

범주형 데이터로 바꿔 보기

PER, Dividend

PER = 주가 / 영업이익

PER 이 작은 거는 영업이익 에 비해 주가가 낮기 떄문이다.

영업이익 이 음수기 때문에 PER 이 음수

영업이익 이 많이 음수일수록 절댓값 은 작아짐

PER 0, (0~2] , 2~4, 4~

Dividend ~-6, -6~-5, [-5~0), 0

주가변동 데이터 이상치 걸러내기 -> 이상치 처리

다른 클러스터에 배정하거나

새로운 거리를 재거나…… 등등