기업개요 데이터 검토 (은수)

형태소 분석 과정

숫자 제외, 맞춤법통일, 기업명 띄어쓰기//사용자 사전 설정, 영어처리, 명사만 추출,

>> 시각화(Wordcloud)

DART 기업개요 데이터 추가

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

주가변동 데이터 검토 (동헌)

(((updown%로 처리했을 때의 문제점.)))

첫날 기준으로, 100%로 설정하기

Dtw\_matrix 만들기 향후 더 적합한 거리계산 metric 선택

>> 시각화(line graph)

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

모델 학습 준비 (도형)

Vectorizer 파라미터, : ngram, min,max\_df , Stop words

Elbow, shilouette, linkage -> 적합한 클러스터 개수 선정: 동헌님 설명

Linkage: (method에 따라 다름: 데이터특성에따라)

DTW 거리 계산

분할적 클러스터링 : Kmeans(n\_cluster = 6) ,timeseriesKmeans, GMM,

밀도기반 클러스터링? : ((DBSCAN, Meanshift)) -> 말로만,,하고 넘어가기?

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

다양한 클러스터링 결과 (다같이 첨언)

기업개요 데이터 시각화 (SVD, PCA, LDA 중 하나)

주가변동 데이터 시각화 (SVD, PCA, LDA 중 하나)

향후 적합한 알고리즘 선택

향후 계획

재무지표 데이터 추가 후, 3번째 대안에 따라 진행할 예정.

“3번째 대안”에 대한 설명