



3대 양념채소 가격 예측

-트렌디한 지표를 활용하여

DATA MINING TEAM

김동영
강수경
김재희
유경민
최윤희



01 주제 선정 배경

“백종원이 양파 볶자 생긴 일?”

폭락했던 양파값이 뛴다

외식사업가 백종원씨가 양파 소비를
촉진하기 위해 유튜브 영상을 올린 뒤

폭락한 양파 가격이 소폭이나마 올랐다.



“TREND”한 변수를 사용해보자!

앞서 말한 일반 변수들이 설명하지 못하는
“급등 급락”과 같은 현상을 설명하고 싶다!

01 기존 가격
예측모델의 한계

02 분석대상 선정
이유

03 도매가격 사용
이유

04 데이터 출처 소개

05 원시데이터 크기
줄이기



01 주제 선정 배경

데이터 출처 소개

- 01 기존 가격
예측모델의 한계
- 02 분석대상 선정
이유
- 03 도매가격 사용
이유
- 04 데이터 출처 소개
- 05 원시데이터 크기
줄이기



농림축산식품
공공데이터 포털
농림축산식품부

“원천 실시간 경락가격 원시데이터”



전국 도매시장에서 거래되는 농산물의 **경매 낙찰 가격**과 **상품 상세 정보**



02 데이터 소개-도매데이터

경매구분코드명

: 도매시장 거래 종류 구분

01 구데이터 변수

02 경매일자

03 경매구분코드

04 부류명

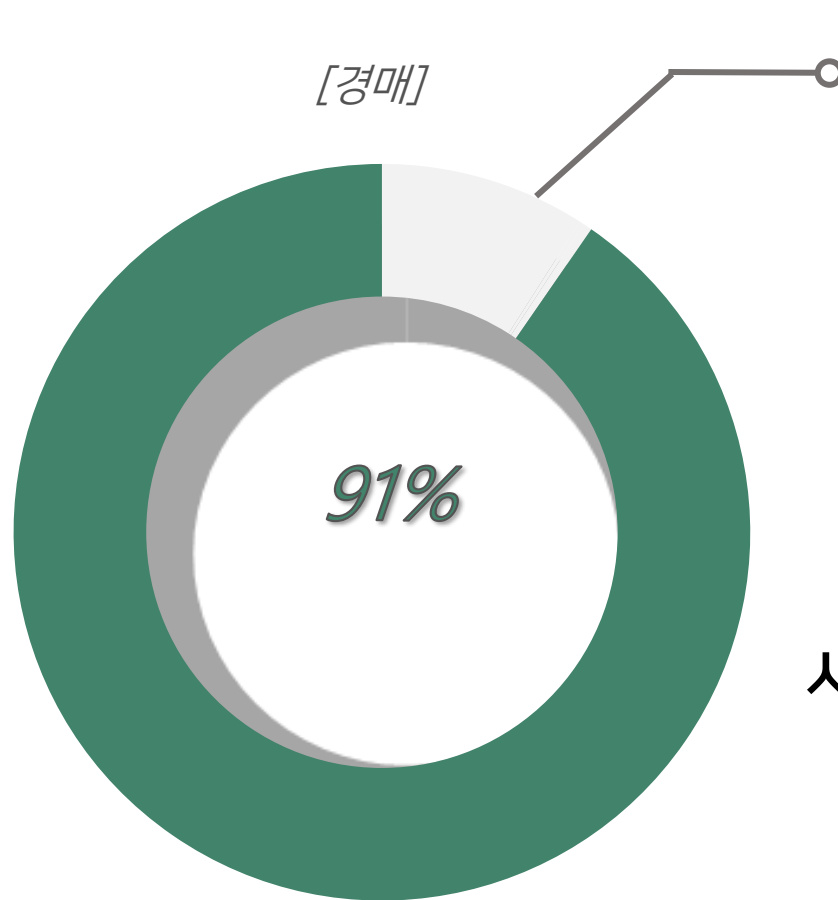
05 거래단량

06 크기명

07 산지명

08 그 외 변수들

09 코드변수



정가수의

정가매매 : 농산물의 **가격**을 미리 정하고 거래

수의매매 : **거래처**를 미리 정하고 거래



시장의 원리에 따라 결정된 가격이 아니므로

정가수의 데이터는 **삭제**



02 데이터 소개-도매데이터

코드 변수

: 코드화된 데이터

01 구데이터 변수

02 경매일자

03 경매구분코드

04 부류명

05 거래단량

06 크기명

07 산지명

08 그 외 변수들

09 코드변수

NAME

법인명

품종명

포장상태명

등급명

출하구분명

산지명

시장명

CODE

도매시장법인코드

품종코드

포장상태코드

등급코드

출하구분코드

산지코드

시장코드

&

PROBLEM

변수별로 코드화 된 변수가 존재하여
데이터의 크기가 커짐



코드북을 만들자!



03 데이터 이상치 처리

이상치처리 – IQR with Log

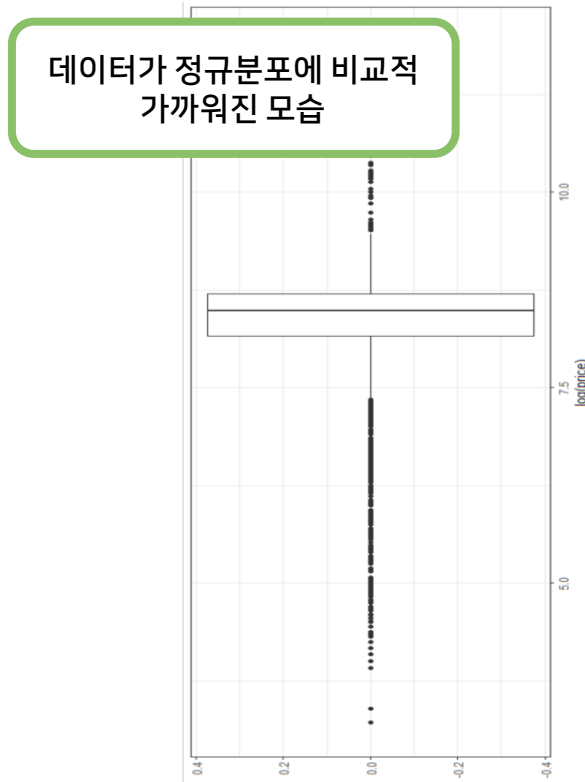
01 가격데이터
분포

02 이상치처리

03 급등급락일
정의

04 볼린저밴드

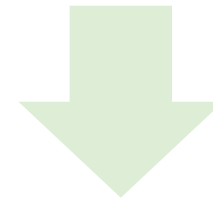
05 급등락일 추출



깎마늘 boxplot

SOLUTION

데이터에 **log** 취하기



- 정규분포에 비슷해짐
- 상한값: 13600원, 하한값: 1550원

➡ 57000개 데이터 중 2000개 제거



변동성 탐지?

“**TREND**”한 변수를 사용해보자!

앞서 말한 일반 변수들이 설명하지 못하는
“**급등 급락**”과 같은 현상을 설명하고 싶다!

HOW?



급변하는 날짜를 찾아서
그날에 **무슨 일**이 일어났는지 찾아보자!

01 가격데이터
분포

02 이상치처리

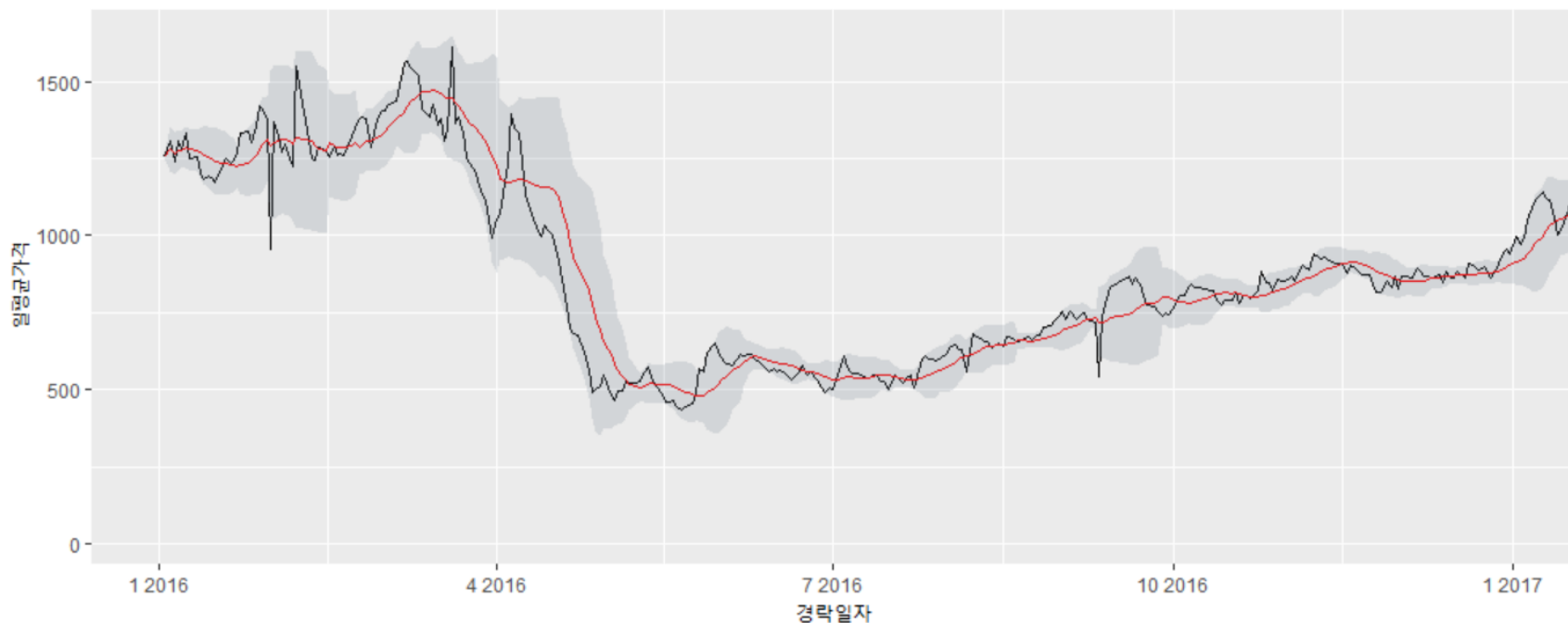
03 급등급락일
정의

04 볼린저밴드

05 급등락일 추출



볼린저 밴드



시계열 데이터에서 이동 평균, 이동 표준편차를 이용하여
데이터가 특정 구역 안에 있을 확률을 시각화한 것

01 가격데이터
분포

02 이상치처리

03 급등급락일
정의

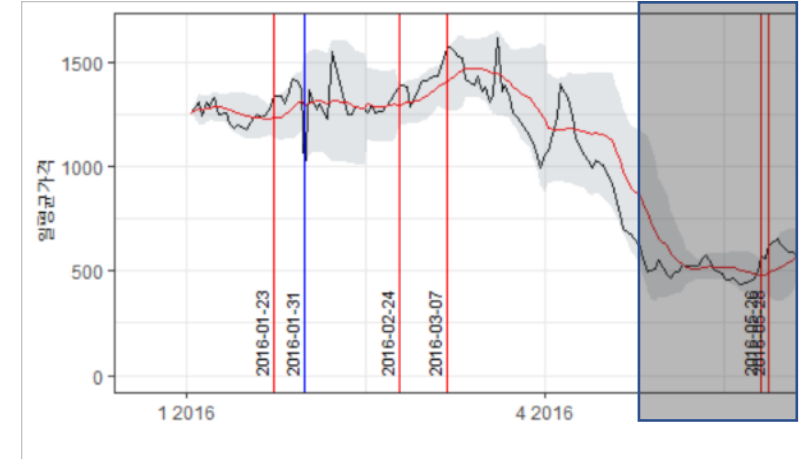
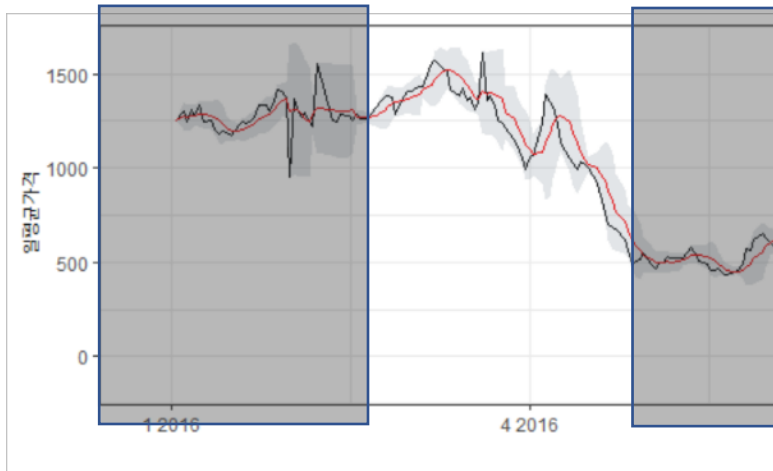
04 볼린저밴드

05 급등락일 추출



04 변동성 탐지

시각화 및 급등락일 추출



탐색일 늘어남



더 과거의 데이터를 탐색



현재 데이터 반영정도 낮음



볼린저 밴드가
현재의 변화에 둔해짐



현재 데이터가 볼린저 밴드
벗어날 가능성이 높아짐



더 많은 급등락일 추출

01 가격데이터
분포

02 이상치처리

03 급등급락일
정의

04 볼린저밴드

05 급등락일 추출



NAVER

01 뉴스 크롤링

02 형태소분석

03 LDA

04 토픽해석

05 다음주예고

마늘

검색

가격 이외에도 다양한 기사를 보기 위해서!

마늘 급등

검색

가격 상승과 관련된 기사를 보기 위해서!
(급등, 폭등 중 급등 기사수가 더 많음)

마늘 폭락

검색

가격 하락과 관련된 기사를 보기 위해서!
(급락, 폭락 중 폭락 기사수가 더 많음)



약 40000건

'마늘 크롤링.csv'
'양파 크롤링.csv'
'고추 크롤링.csv'



약 2000건

'마늘 급등 크롤링.csv'
'양파 급등 크롤링.csv'
'고추 급등 크롤링.csv'



약 1000건

'마늘 폭락 크롤링.csv'
'양파 폭락 크롤링.csv'
'고추 폭락 크롤링.csv'



01 뉴스 크롤링

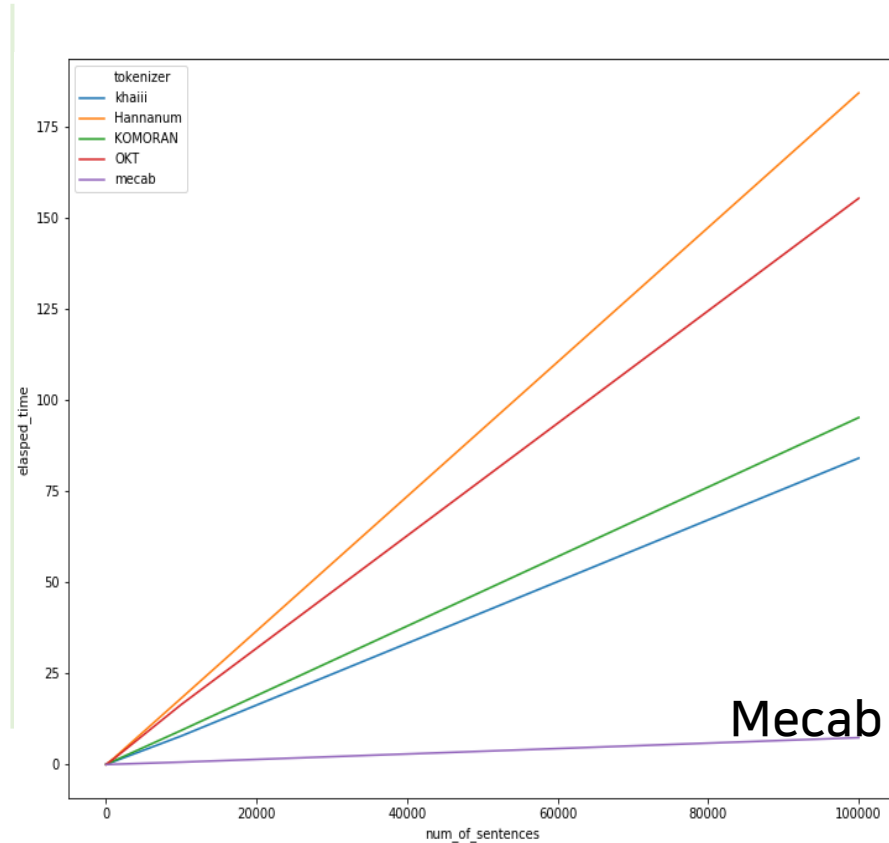
02 형태소분석

03 LDA

04 토픽해석

05 다음주예고

Mecab



형태소 분석기 Mecab

- 일본어 형태소 분석 방법 바탕
- 압도적인 속도
- 보편적인 성능

뉴스기사의 특성

- 기사개수 많음
- 정제된 데이터

→ 속도·성능 모두 Mecab으로 가능



토픽모델링

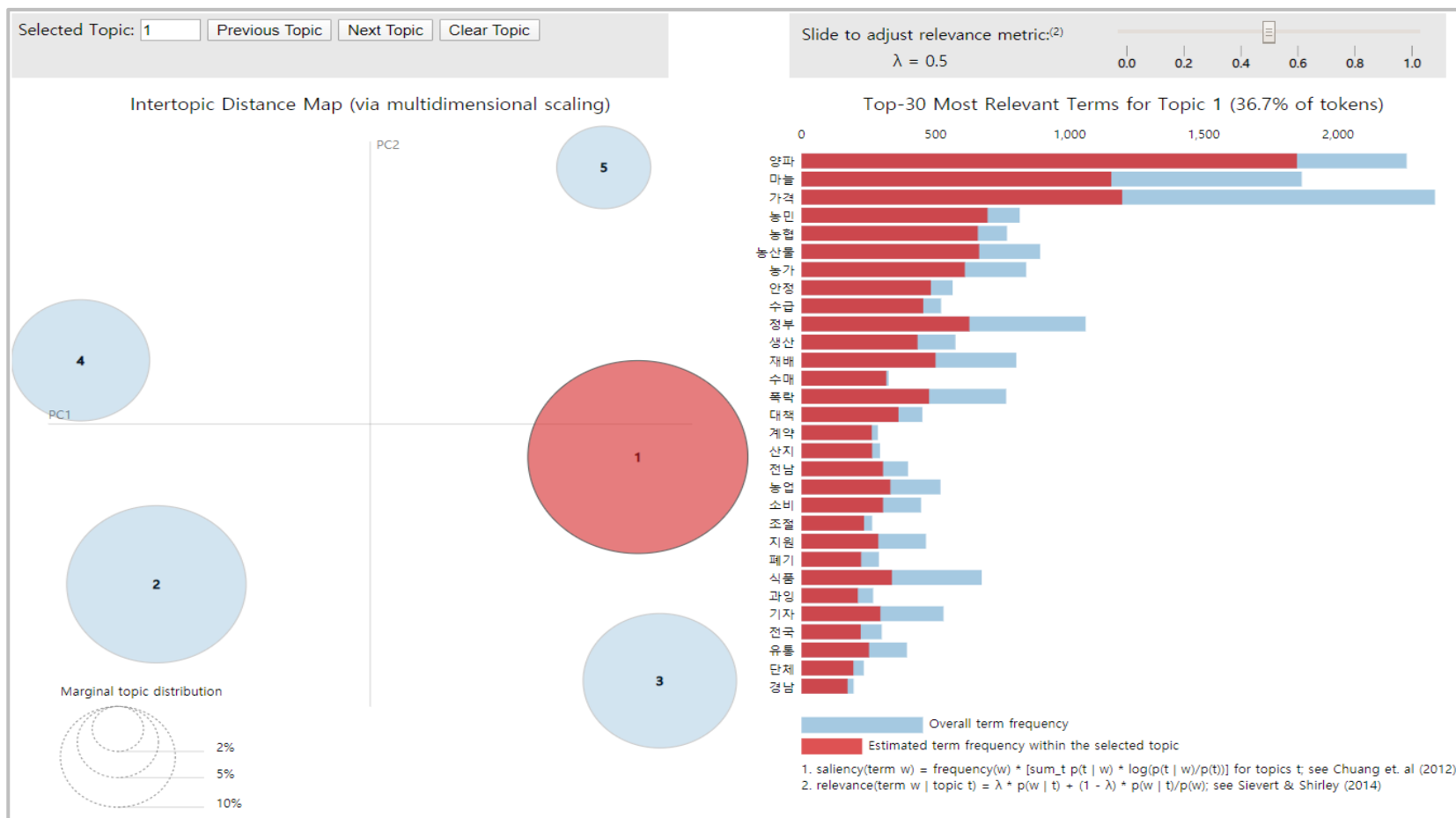
01 뉴스 크롤링

02 형태소분석

03 LDA

04 토픽해석

05 다음주예고



자세히 살펴볼까요?



"급등" 관련 기사

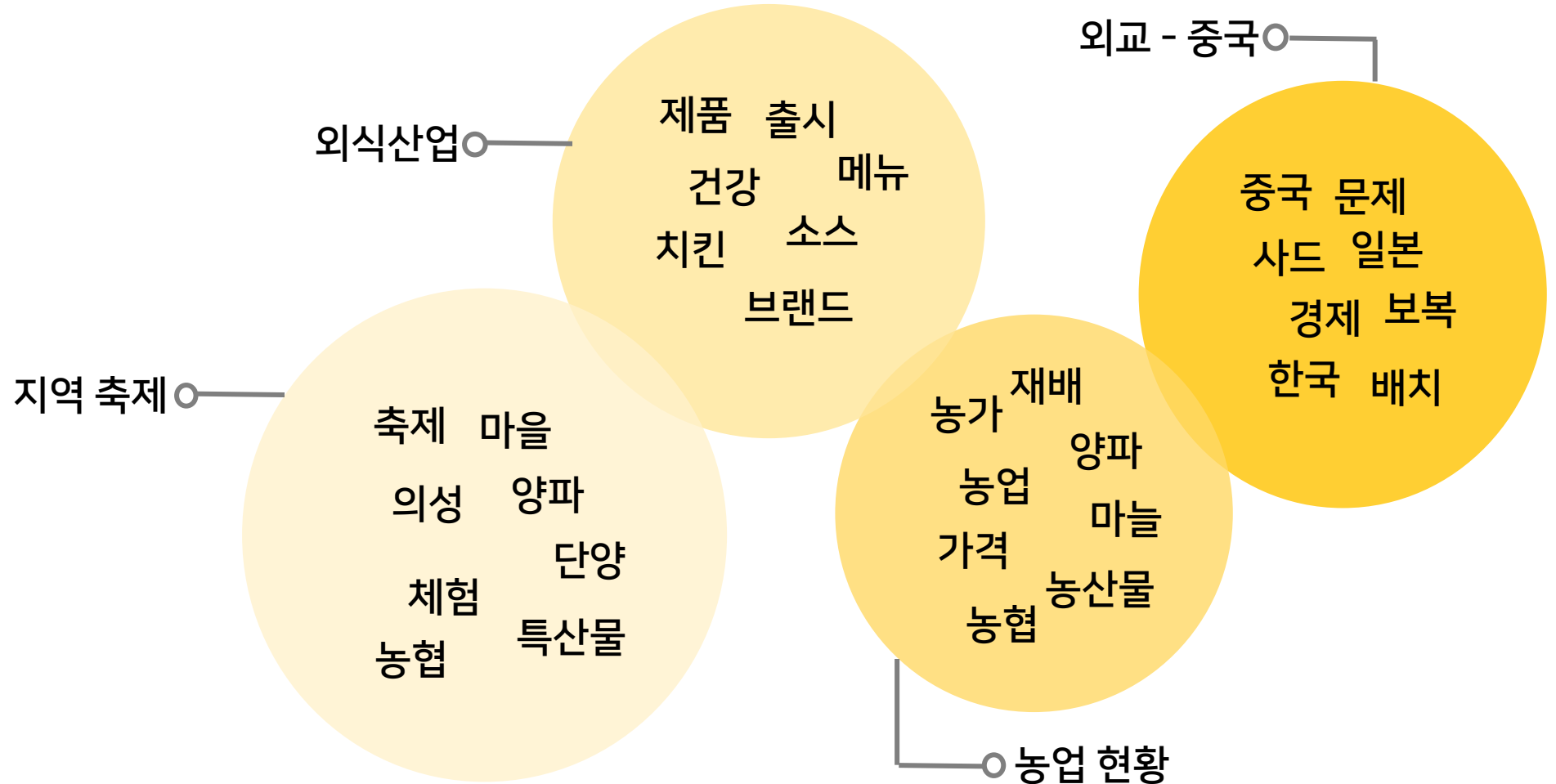
01 뉴스 크롤링

02 형태소분석

03 LDA

04 토픽해석

05 다음주예고





파생변수생성

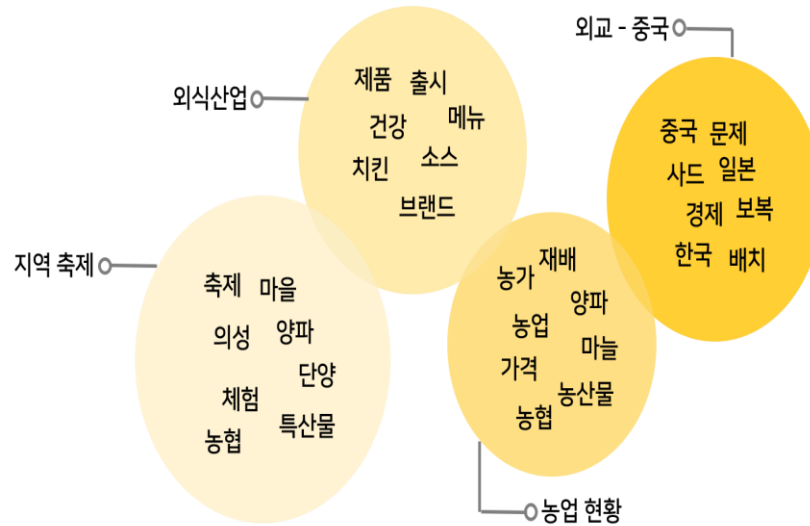
01 뉴스 크롤링

02 형태소분석

03 LDA

04 토픽해석

05 다음주예고



각 **토픽**들에 대한 특성 파악

특성을 활용한 **파생변수** 생성



06 다음주예고

Train & Test



01 뉴스 크롤링

02 형태소분석

03 LDA

04 토픽해석

05 다음주예고

Train
(16~18)

Test
(19)