

텍스트 마이닝을 활용한 금통위 의사록 분석

Deciphering Monetary Policy Board Minutes through Text Mining Approach : The Case of
Korea

Team 3

김형석 송지혜 이진아 이하영 홍현택

CONTENTS

I. 프로젝트 배경

II. 프로젝트 과정

III. 프로젝트 결과

IV. 프로젝트 시사점

V. 프로젝트 정리

I. 프로젝트 배경



1.1 프로젝트 개요

< 텍스트 마이닝을 활용한 금통위 의사록 분석 >



텍스트 마이닝



기준금리 예측

1.1 프로젝트 개요



통화정책은 적시성과 신뢰성이 중요하다.

그러나 정책 결정자의 주관이 개입되기 때문에 언제나 불확실성을 내포하기도 한다



기준금리 관련 이주열 한국은행 총재 발언

8월31일 금융통화위원회 후 간담회
"향후 성장, 물가 흐름, 금융안정 상황을 더 짚어보겠다.
경기도 보지만 금융안정 상황에 더 유의할 필요가 있다"

9월27일 미 금리 인상 후 출근길
"금리결정에는 거시변수가 제일 중요하고, 저금리가
오래갔을 때 금융불균형이 어느 정도 쌓일지 종합적으로
고려해 최적의 결정을 할 것"

10월4일 경제동향간담회 모두발언
"금융불균형이 누증되고 있다. 금융불균형을 점진적으로
해소하는 등 거시경제를 안정적으로 운영해야 한다"

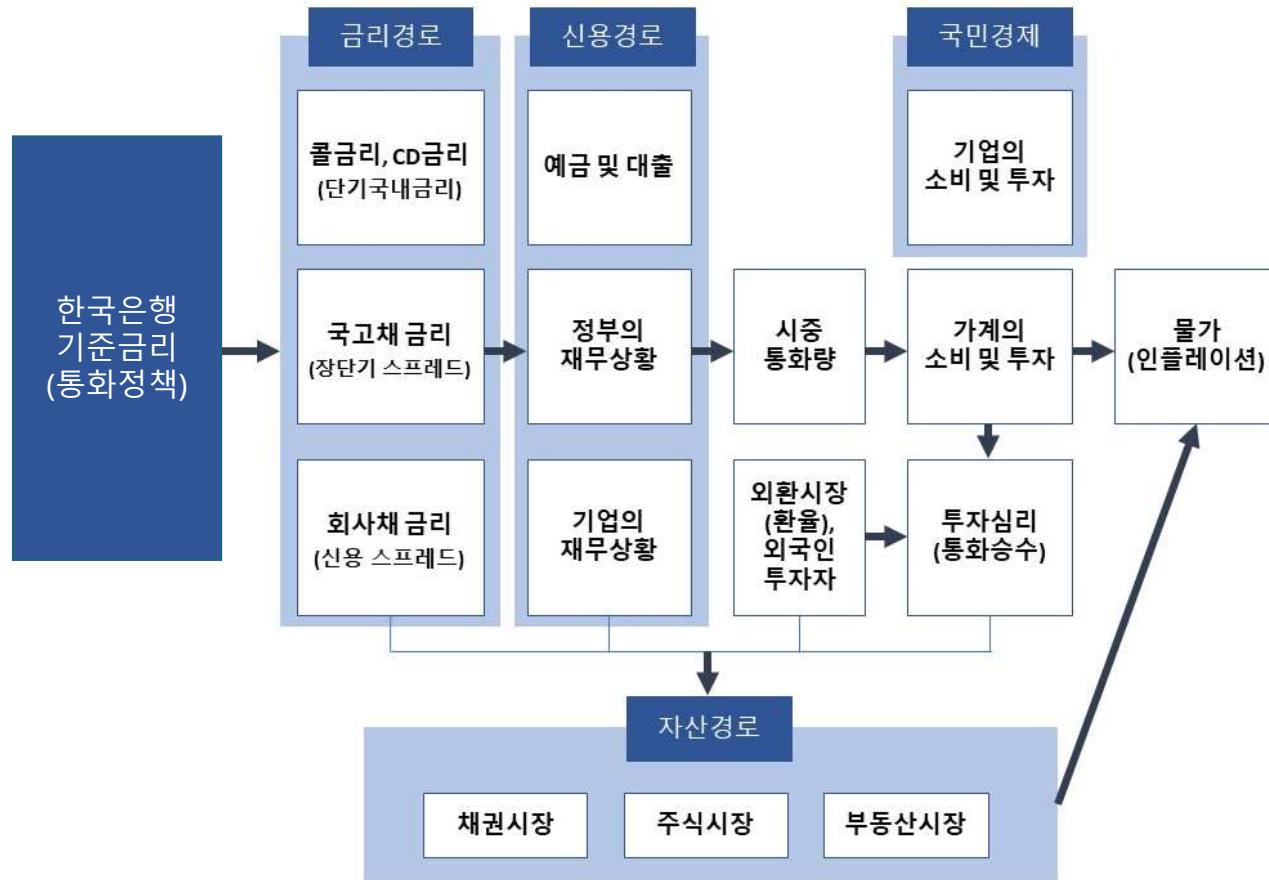
10월5일 기자단 워크숍
"잠재성장을 수준의 성장세가 이어지고, 물가 목표 수준에
점차 근접해나간다는 판단이 선다면 금융안정도 비중 있게
고려해야 할 시점이다. 외부 의견을 의식해서 금리인상이
적절치 않은데도 하는 결정은 하지 않을 것"

출처 : 세계일보

page

1.1 프로젝트 개요

통화정책 파급경로



1.2 프로젝트 목적

< 프로젝트 목적 >



1. 데이터 처리부터 분석까지 논문을 충실히 따르며 직접 구현
2. 프로젝트 결과물의 활용 및 응용 가능성 검토

1.3 데이터 수집 계획, 역할 분배

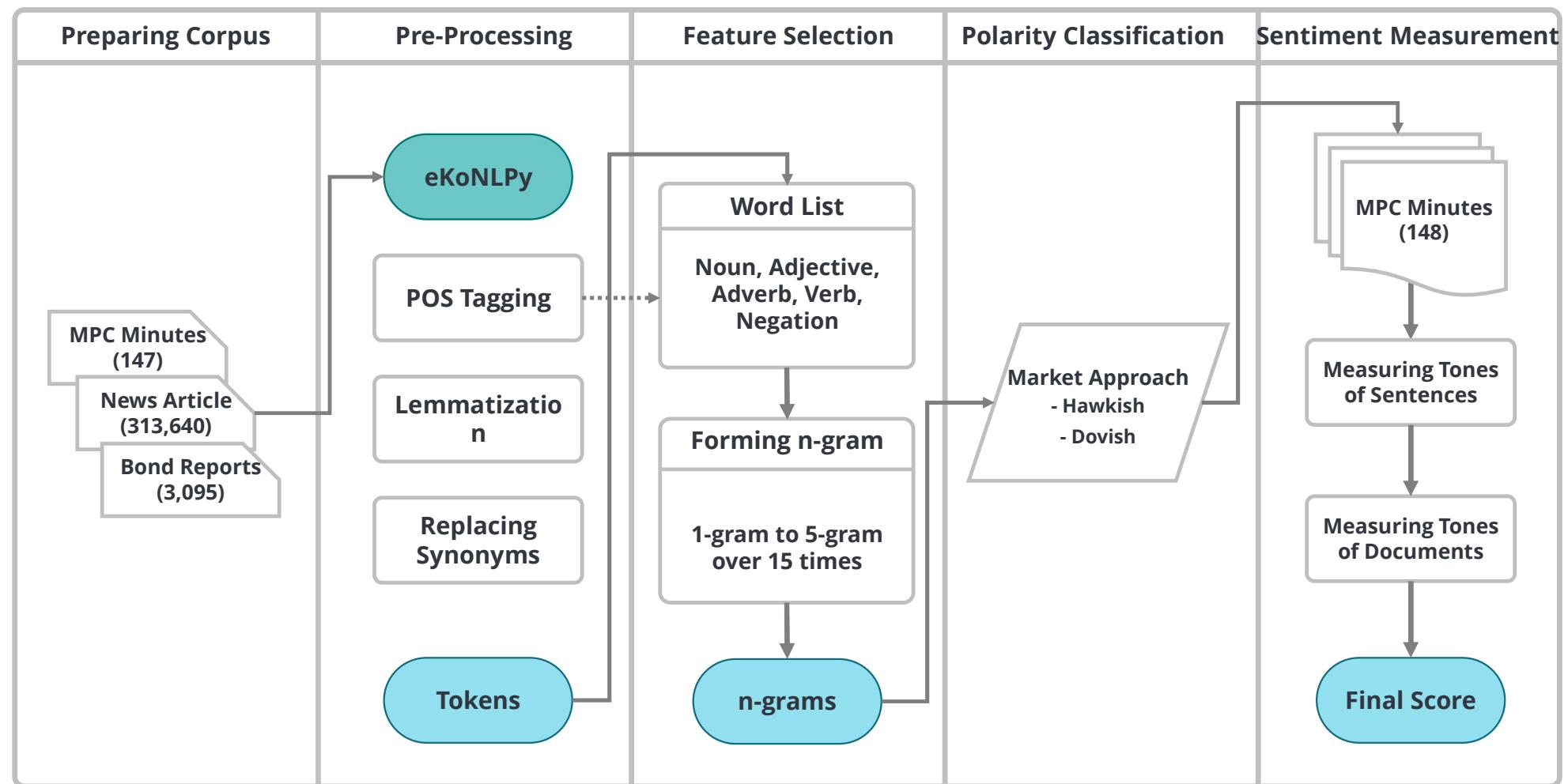
Schedule

구분	일정				
	5/5~5/11	5/12~5/18	5/19~5/25	5/26~6/1	6/2~6/3
논문분석					
데이터 수집					
데이터 전처리					
모델링					
시각화					
발표자료					

II. 프로젝트과정



2. Procedure of Sentiment Analysis



2.1 Preparing Corpus

데이터 수집 과정

01

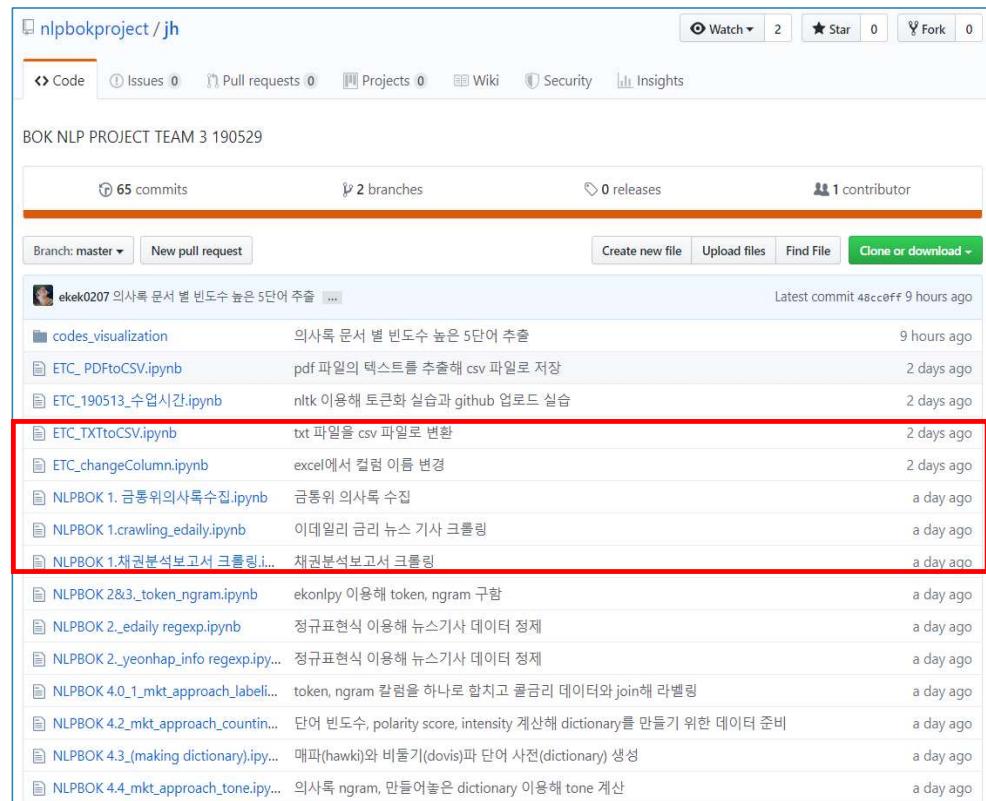
데이터 수집

selenium, BeautifulSoup 활용

02

데이터 저장

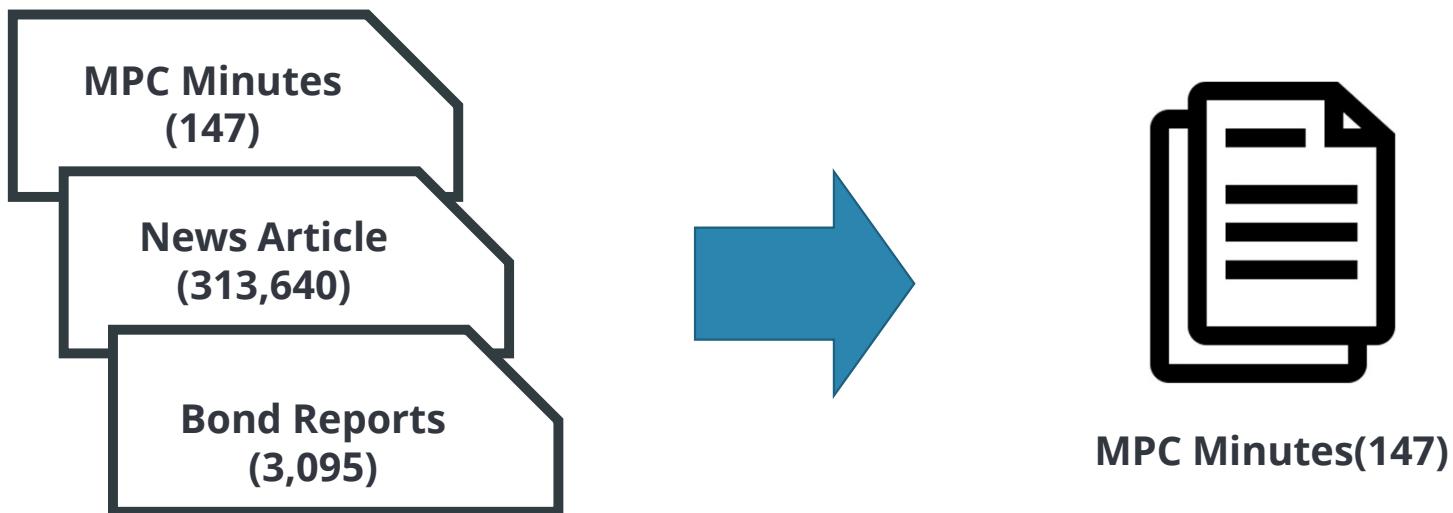
Data Frame 형태로 변환 후
corpus별 csv 형식으로 저장



2.1 Preparing Corpus

데이터 수집 과정

316,882건의 수집된 문서로 Filed-specific-lexicon 구성



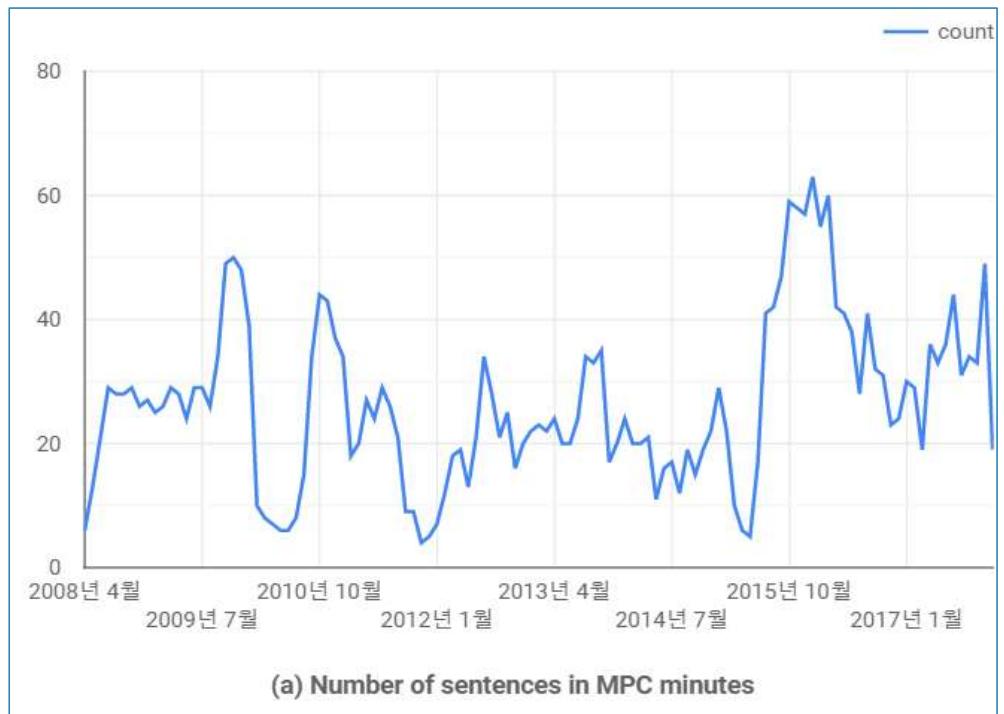
2.1 Preparing Corpus

금통위 의사록(MPB minutes)

활용 Data : 금융통화의원회 의사록

- 수집 출처 : 한국은행
- 기간 범위 : 2005년~2017년(13개년 데이터)
- 데이터 수 : MPB minutes 147건

date	Economic Situation count	Foreign Currency count	Financial Markets count	Monetary Policy count	Participant Views count	Government View count	Foreign Currency ngram	Financial Markets ngram
0 20060810	0	0	0	0	78	0	NaN	NaN
1 20180412	74	69	0	0	400	0	fed/NNG,금리/NNG,인상/NNG,속도/NNG,가속/NNG,국제/NNG,금융 시...	NaN
2 20070608	0	0	0	0	80	0	NaN	NaN
3 20070111	0	0	0	0	96	0	NaN	NaN
4 20051208	2	0	0	0	84	0	NaN	NaN
5 20170223	46	51	0	0	362	0	경제/NNG,성장률/NNG,잠재/NNG,성장률/NNG,상회/NNG,국제/NNG,금융...	NaN
6 20120209	3	0	0	0	194	0	NaN	NaN
7 20121109	25	12	15	0	318	0	외화차입/NNG,여건/NNG,양호/NNG, 글로벌/NNG,유동성/NNG,증기/NNG,...	경기/NNG,하방/NNG,위험/NNG,증대/NNG,금융시장/NNG,변동성/NNG,안...
8 20180524	88	73	0	0	384	0	가계/NNG,대출/NNG,권리/NNG,강화/NNG,가계/NNG,대출/NNG,증기/N...	NaN
9 20060112	0	0	0	0	69	0	NaN	NaN
10 20181018	79	59	0	0	441	0	단기/NNG,금융시장/NNG,변동성/NNG,확대/NNG,fed/NNG,금리/NNG...	NaN
11 20090910	0	0	0	0	10	0	NaN	NaN
12 20140410	43	12	22	0	362	0	유가/NNG,원자재/NNG,가격/NNG,하락/NNG,기대/NNG,시장/NNG,크리...	양적완화/NNG,죽소/NNG,금리/NNG,인상/NNG,가계/NNG,보체/NNG,문



2.1 Preparing Corpus

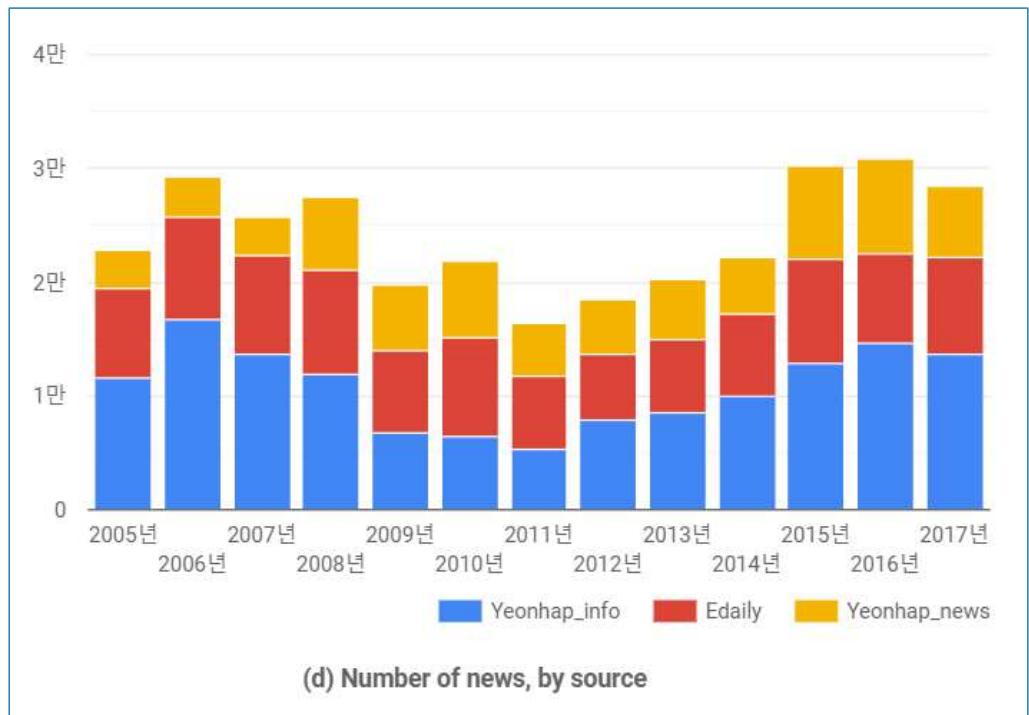
뉴스 기사(News Article)

활용 Data : 뉴스 데이터 (키워드 : '금리')

- 수집 출처 : 이데일리, 연합뉴스, 연합인포맥스
- 기간 범위 : 2005년~2017년(13개년 데이터)
- 데이터 수 : 313,640건

종류	연합뉴스	연합인포	이데일리	총 합계
문서 수	72,149	139,974	101,517	313,640

	title	date	url	script
NaN	title	date	url	script
NaN	title	date	url	script
NaN	title	date	url	script
0.0	인도 모디 총리 신년사서 '화폐개혁' 정당성 주장	2017-01-01	http://www.edaily.co.kr/news/newspath.asp?news...	[이데일리 방성훈 기자] 인도의 모디 총리가 '부과, 같은 돈, 위조 지폐는 인도의 ...
0.0	"조율확실성의 시대"...원난한 금리정책 예고	2017-01-01	http://www.edaily.co.kr/news/newspath.asp?news...	이주얼 한국은행 총재가 지난달 15일 서울 종로 한은 본관에서 열린 금융통화위원회 ...
0.0	"조율확실성의 시대"...원난한 금리정책 예고	2017-01-01	http://www.edaily.co.kr/news/newspath.asp?news...	이주얼 한국은행 총재가 지난달 15일 서울 종로 한은 본관에서 열린 금융통화위원회 ...
0.0	[새해 부동산 재테크]"금리 변수...부동산 투자는 본 수석 접근 필요"	2017-01-01	http://www.edaily.co.kr/news/newspath.asp?news...	[이데일리 박태진 기자] 새해 부동산시장 전망은 그야말로 '안갯 속'이다. 원탁 다양 ...
0.0	기업운영, '맑은 해'적금은 꼭이오 이벤트 실시	2017-01-01	http://www.edaily.co.kr/news/newspath.asp?news...	[이데일리 김경운 기자] IBK기업은행은 새학기를 맞아 오는 2월부터 2월말까지 영업 ...
0.0	[신년사]박정원 두산그룹 회장 "재무건전성 강화해 풀확실성 확보"	2017-01-01	http://www.edaily.co.kr/news/newspath.asp?news...	박정원 두산그룹 회장[이데일리 성문재 기자] 박정원 두산 (000150)그룹 회장이 ...
0.0	[신년사]정지원 증권금융 사장 "위기대응 위해 혁신...도전해야"	2017-01-01	http://www.edaily.co.kr/news/newspath.asp?news...	[이데일리 유재희 기자] 정지원 한국증권금융 사장(사진)은 새해를 확실한 경영환경에 ...



2.1 Preparing Corpus

채권 분석 보고서(Bond Analysts' Reports)

활용 Data : 채권 분석 보고서)

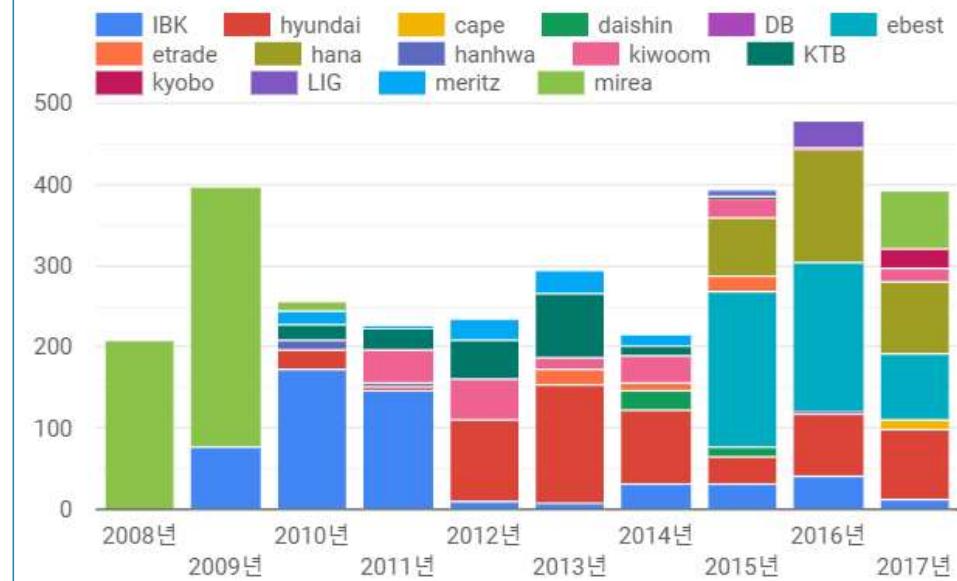
- 수집 출처 : 네이버 금융 > 투자전략
- 기간 범위 : 2008년~2017년(10개년 데이터)
- 데이터 수 : 15개 금융사 3,095건

df.head()

	date	securities	title	report	tokens	ngram
0	13.10.17	KTB투자증권	10월 정치 불확실성 미완의 종결	10월 정치 불확실성 미완의 종결 17 Oct. 2013 미 상원 부채한도 협상...	정치/NNG, 불확실성/NNG, 미완/NNG, 종... 결/NNG, 미/NNG, 부채한도/NNG, 협...	정치/NNG, 불확실성/NNG, 해소/NNG, 채권...
1	10.12.09	KTB투자증권	뉴트럴 엔딩 Neutral Ending	9 December 2010 2010년 12월 금융통화위원회 뉴트럴 엔딩 (Neut...	뉴트럴/NNG, 엔딩/NNG, 뉴트럴/NNG, 통화...	재정/NNG, 문제/NNG, 지정학/NNG, 위...
2	15.01.19	키움증권	저금리 장기화 주요 수출국과의 느슨해진 관계	Microsoft Word - 150119-WeeklyBondMarket.doc...	저금리/NNG, 장기/NNG, 수출국/NNG, 느...	글로벌/NNG, 안전자산/NNG, 선호/NNG, 강...
3	15.12.30	이베스트투자증권	금통위 의사록 금리 인하 보다 구조개혁 중요	Credit 손소현 02-3779-0055 thecredit@ebestsec.co...	손/NNG, 전병/NNG, 하/XSV, 증가/NNG, 금리/NNG, 스프레드/NNG, 금리/N...	경제/NNG, 성장률/NNG, 전망/NNG, 하향/NNG, 성장/NNG, 하방/NNG, 위험...
4	15.02.09	현대차증권	전화위복	<그림1> 주요 여전체 신용등급 변동 현황 01234567812.1 13.1 14....	여전체/NNG, 신용등급/NNG, 변동/NNG, 현황/NNG, 무렵/NNG, 자료/NNG, 전...	외국인/NNG, 국채선물/NNG, 매도/NNG, 금리/NNG, 상승/NNG, 우려/NNG, 완...

df.head()['ngram'][0]

'정치/NNG:불확실성/NNG:해소/NNG:채권시장/NNG:금리/NNG:상승/NNG:디폴트/NNG:가능성/NNG:우려/NNG:채권시장/NNG:약세/NNG:되돌/YY,높/VN:정치/NNG:
불확실성/NNG:기간스프레드/NNG:확대/NNG:안전자산/NNG:우호적/YAX,단기/NNG:불확실성/NNG'



(f) Number of bond analysts' reports, by section

2.2 Pre-Processing

데이터 전처리 과정

01

Pre-processing

- Tokenization
- Normalization

02

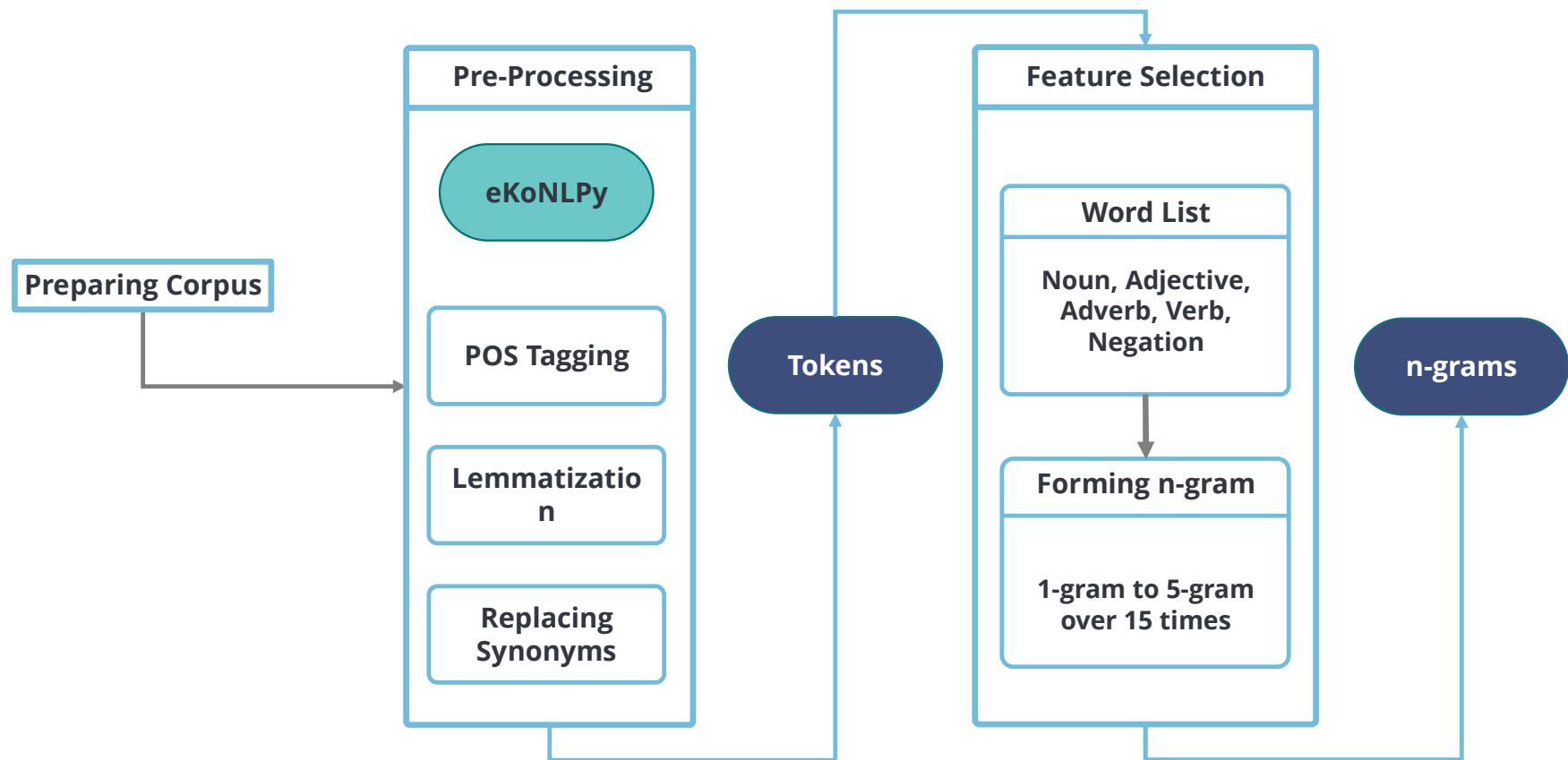
Feature Selection

- Token → n-gram

nlpbokproject / jh		
Code	Issues 0	Pull requests 0
Projects 0	Wiki	Security
BOK NLP PROJECT TEAM 3 190529	65 commits	2 branches
Branch: master	New pull request	Create new file Upload files Find File Clone or download +
ekek0207 의사록 문서 별 빈도수 높은 5단어 추출	9 hours ago	Latest commit 48cc0ff 9 hours ago
codes_visualization 의사록 문서 별 빈도수 높은 5단어 추출	9 hours ago	
ETC_PDFtoCSV.ipynb pdf 파일의 텍스트를 추출해 csv 파일로 저장	2 days ago	
ETC_190513_수업시간.ipynb nltk 이용해 토큰화 실습과 github 업로드 실습	2 days ago	
ETC_TXTtoCSV.ipynb txt 파일을 csv 파일로 변환	2 days ago	
ETC_changeColumn.ipynb excel에서 컬럼 이름 변경	2 days ago	
NLPBOK 1. 금통위의사록수집.ipynb 금통위 의사록 수집	a day ago	
NLPBOK 1.crawling_edaily.ipynb 이데일리 금리 뉴스 기사 크롤링	a day ago	
NLPBOK 1.채권분석보고서 크롤링.ip... 채권분석보고서 크롤링	a day ago	
NLPBOK 2&3._token_ngram.ipynb ekonlpy 이용해 token, ngram 구함	a day ago	
NLPBOK 2._edaily regexp.ipynb 정규표현식 이용해 뉴스기사 데이터 정제	a day ago	
NLPBOK 2._yeonhap_info regexp.ip... 정규표현식 이용해 뉴스기사 데이터 정제	a day ago	
NLPBOK 4.0_1_mkt_approach_labeli... token, ngram 칼럼을 하나로 합치고 콜금리 데이터와 join해 라벨링	a day ago	
NLPBOK 4.2_mkt_approach_countin... 단어 빈도수, polarity score, intensity 계산해 dictionary를 만들기 위한 데이터 준비	a day ago	
NLPBOK 4.3_(making dictionary).ipy... 매파(hawk)와 비둘기(dovis)파 단어 사전(dictionary) 생성	a day ago	
NLPBOK 4.4_mkt_approach_tone.ip... 의사록 ngram, 만들어놓은 dictionary 이용해 tone 계산	a day ago	

2.2 Pre-Processing

데이터 전처리 과정



2.2 Pre-Processing

Tokenization

```

news = []
mecab = Mecab()
for i in range(1,2): #1월부터 12월까지 돌림
    df = pd.read_csv("./news_201"+str(i)+"_preprocessed.csv")
    #print(news)

    #클러온 csv파일의 각 인덱스 script를 토큰화
    #for j in range(df.shape[0]): #index 수만큼 range설정
    for j in range(0,df.shape[0]): #index 수만큼 range설정
        #print(j)
        news.append(df['script']) #script만 토큰화하기 위해 script 선택

    t=mecab.pos(news[j]) #j번째 script의 tokenize
    t=str(t) #tokenize된 list 데이터를 string화
    #tstr = ''.join(t)

    df.loc[j, 'token'] = t #token 열 추가해서 j index에 해당하는 부분에 tokenize된 데이터 추가.
    #print(df)

df.to_csv("./news_" + str(i).zfill(2) + "-processingCompleted.csv", mode='w')

```

com	date	title	url	script	token	ngram
0	Edaily 2017-01-01	"조불확실성의 시대"... 험난한 금리정책 예고	http://www.edaily.co.kr/news/newspath.asp?news...	이주열 한국은행 총재가 지난달 15일 서울 종로 학술 분관에서 열린 금융통화위원회...	본관/NNG 열/VV 참석/NNG 생각/NNG 참석/NNG 높/VA, 의회 거/NV.상승기간/NNG 통화정책...	변동성/NNG 확대/NNG 가능성/NNG 높/VA, 의회 시장/NNG, 변동성/NNG, 확...
1	Edaily 2017-01-01	[새해 부동산 재테크] "금리 변수, 부동산 투자는 보수적 접근 필요"	http://www.edaily.co.kr/news/newspath.asp?news...	새해 부동산시장 전망은 그 아말로 '안갯속'이다. 위낙 다양한 이유와 변수들이 산재...	새해/NNG, 부동산/NNG, 전망/NNNG, 그야말로/MAG, 안갯속/NNG, 위험/MAG...	주택/NNG, 공급/NNG, 과잉/NNG, 우려/NNG, 임대/NNG, 수요/NNG, 확보/N...
2	Edaily 2017-01-01	[신년사]박정원 두산그룹 회장 "재무건전성 강화해 불확실성 극복"	http://www.edaily.co.kr/news/newspath.asp?news...	박정원 두산그룹 회장 박정원 두산(000160)그룹 회장이 새해 재무건전성 강화에...	새해/NNG, 재무/NNG, 정성/NNG, 강화/NNG, 집중/NNG, 하/HSV, 의지/NN...	경영/NNG, 환경/NNG, 불확실성/NNG, 금리/NNG, 인상/NNG, 정책/NNG, 확대...

Normalization

```

#pattern = r'[a-zA-Z0-9_.%+-]+@[a-zA-Z0-9_.%+-]+\.' #이메일자우기
#pattern = '&lt;a[a-zA-Z_.%+-]&gt;.' #이메일자우기
#pattern = r'[0-9]+/[0-9]+/[0-9]+.' #날짜 자우기
#pattern = r'[0-9]+:[0-9]+.' #시간 자우기
#pattern = '출고'
#pattern = '[가-힣]+ [가-힣]+'
#pattern = '[[가-힣][a-zA-Z0-9_.%+-]+]' # 관련기사 ◀+▶ 관련기사 ◀+▶ 뒤의 관련기사 제목 제거
pattern = '▶ 관련기사 ◀+' #▶ 관련기사 ◀+▶ 뒤의 관련기사 제목 제거
pattern = '사진=[가-힝]+[[s] 제공]' # 사진=○○○ 제공 제거
pattern = '#[[이데일리][■■■가-일]+#[#이데일리]]' 제거
pattern = '#[[사진=(가-힝)]+[s] 기자]' #[사진=○○○ 기자] 제거
pattern = '#[[이데일리][s] [가-힝]+[s] [가-힝]+#[#이데일리 ~~~]]' 제거
pattern = '#[[이데일리][s] [가-힝]+[s] 기자]' #[이데일리 ○○○ 기자] 제거
pattern = '#[[^a-zA-Z0-9_+]]' # /] 안의 기타 문자 제거
pattern = '[가-힝]+[s]+기자'
pattern = '[가-힝]+ 특파원'
pattern = 'r'[a-zA-Z0-9_.%+-]+@[a-zA-Z0-9_.%+-]+\.' #이메일자우기

# #//^a-zA-Z0-9_+#
error_ls=[]

for file in files:
    df = pd.DataFrame(columns=['com', 'date', 'title','url','script'])
    df = pd.read_csv(file, names=['com', 'date', 'title','url','script'])

    script_ls = df['script']

    new_script_ls = []
    for i in script_ls:
        mail_ls = []
        try:
            mail_ls = re.findall(pattern, i)
            print(mail_ls)
        except:
            print(file)
        if mail_ls != []:
            for j in mail_ls:
                new_script = i.replace(j, '')
            new_script_ls.append(new_script)
        else:
            new_script_ls.append(i)

```

2.2 Pre-Processing

Feature Selection

Out[0]:									
	date	rate	difference	label	media	title	url	news	tokens
0	2012.01.01	3.29	0.03	1	연합 뉴스	2012 예산 확정...경기 위축 대비 조기집행	https://news.naver.com/main/read.nhn? mode=LSD&...	재 정/NNG, 적 자/NNG, 제정적자 14 조 전망... 2013 균형재 정 '순항' 31 일 확정된 2012년 예... 항/NNG, 순 항/NNG, 확 정/NNG, 되/...	In [0]: row['neran']
1	2012.01.01	3.29	0.03	1	연합 뉴스	2012 예산 확정...복지 늘리고 FTA 지원 강화	https://news.naver.com/main/read.nhn? mode=LSD&...	일자 리/NNG, 한미 fta/NNG, 지 원/NNG, 증 액/NNG, 복 지/NNG, 지출비 증/NN...	Out [0]:

2.3 Polarity Classification

극성 분류

01

n-gram 라벨링

- 콜금리 데이터를 통한 라벨링
- 날짜 / n-gram / 라벨링(dovish, hawkish)

02

Naïve Bayes Classifier 모델링

- NBC모델을 기반으로 조건부확률 도출
- 극성점수 도출

03

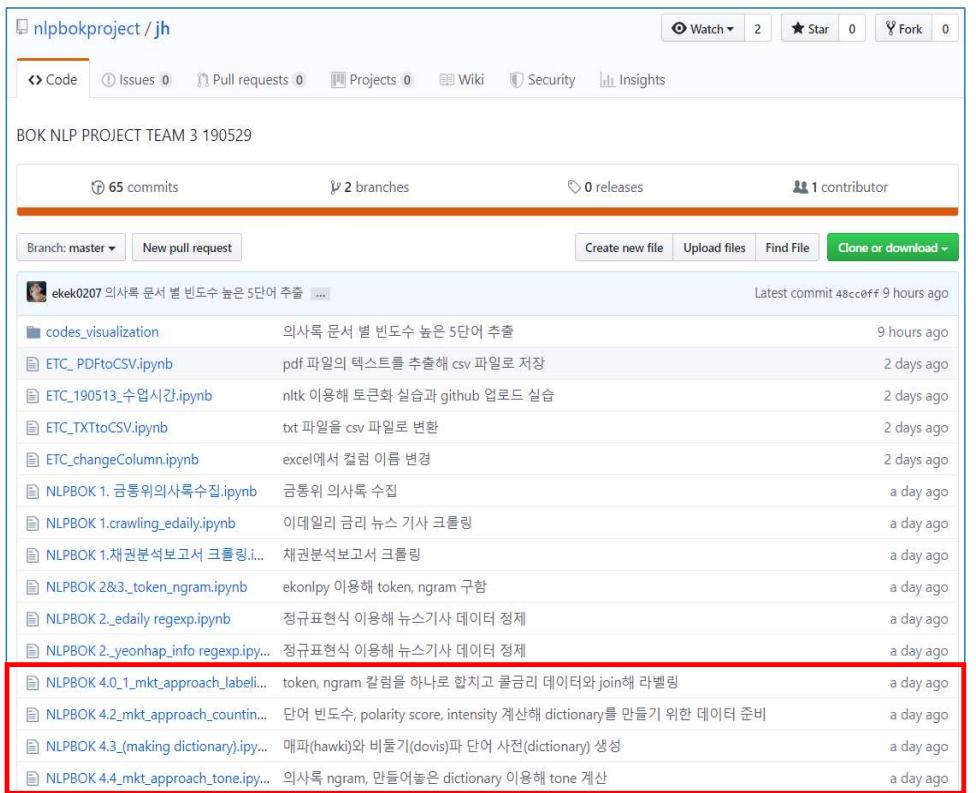
Dictionary

- Dovish / Hawkish n-gram 사전

04

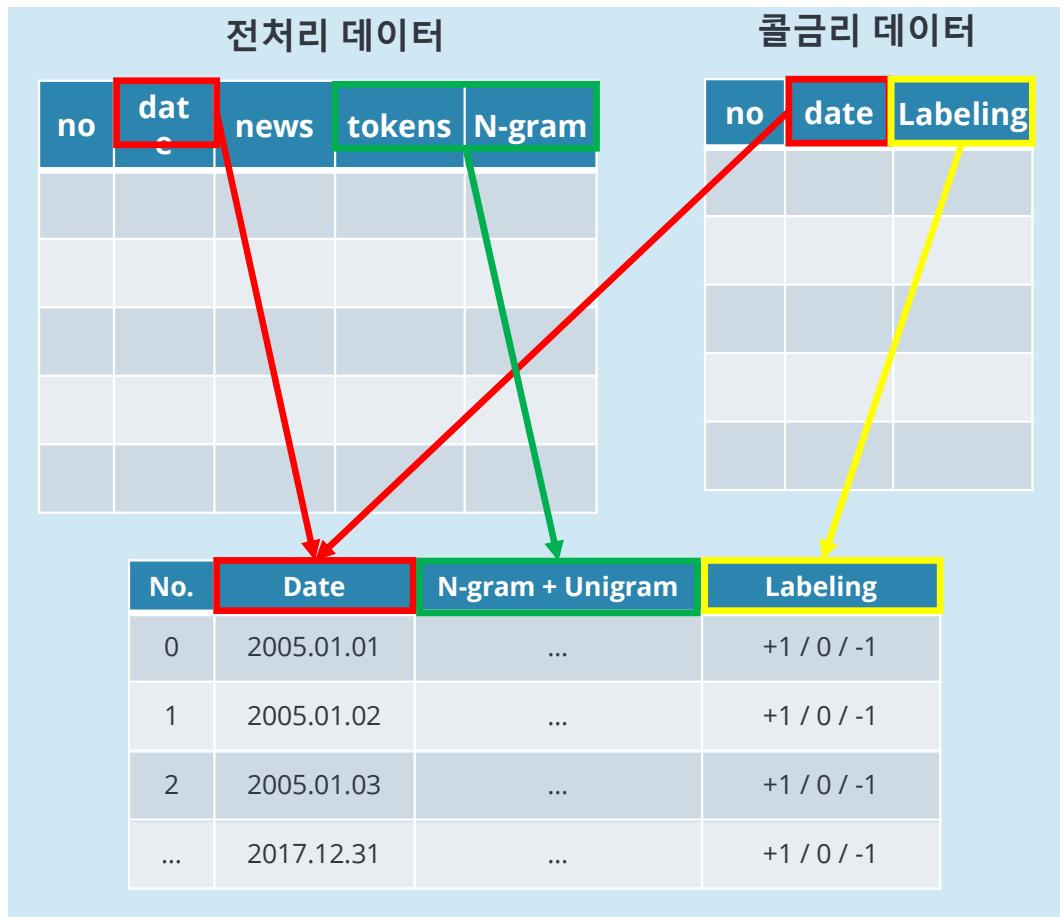
Measuring Sentiments

- Dovish / Hawkish 사전



3.1 Polarity Classification

Step 1. n-gram 라벨링



추출한 token과 n-gram을 콜금리 데이터로 라벨링

- ▶ 키 값은 'date'
- ▶ 콜금리가 상승할 경우 positive(hawkish) : +1
- ▶ 콜금리가 하락할 경우 negative(dovish) : -1

같은 날짜 = 같은 라벨

2.3 Polarity Classification

Step 2. Naïve Bayes Classifier 모델링

기계학습(Machine-learning)중 하나인 단순화를 분류자, Naïve-Bayes Classifier(NBC)를 사용

$$\text{polarity score} = \frac{p(\text{feature} | \text{hawkish})}{p(\text{feature} | \text{dovish})} = \frac{p(\text{feature} | \text{hawkish})/p(\text{hawkish})}{p(\text{feature} | \text{dovish})/p(\text{dovish})}$$

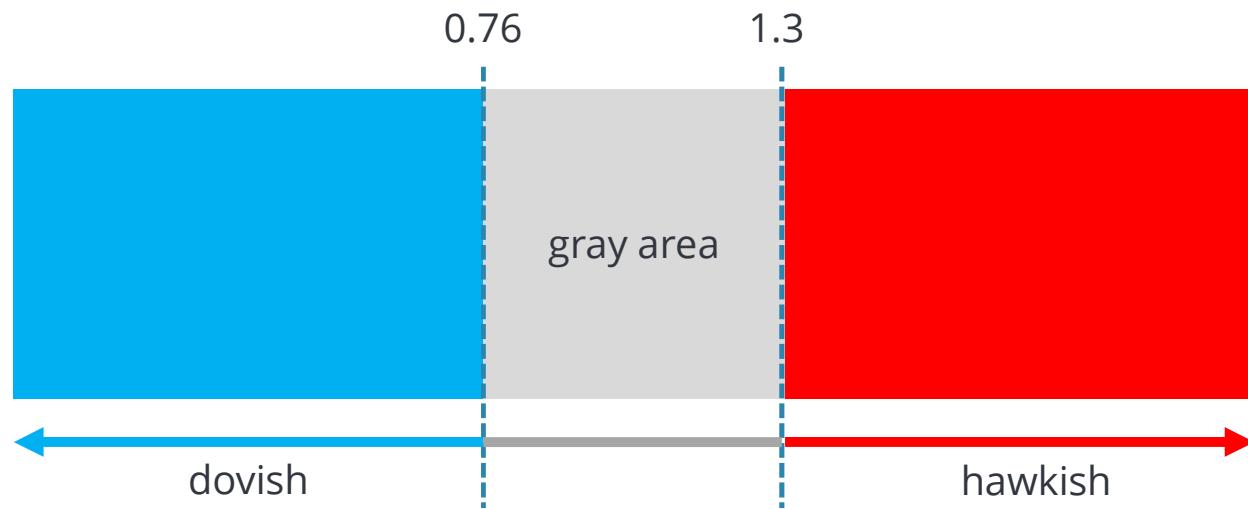
The diagram illustrates the process of calculating polarity scores. On the left, two boxes represent document types: 'Hawkish Documents' and 'Dovish Documents'. Arrows from these boxes point to a central table. A large red arrow labeled '극성점수 도출' (Calculate Polarity Score) points from the table back towards the document boxes.

N-gram + unigram	Counting		Polarit y Score	Intensit y	Label
	Hawkish	Dovish			
인상 /NNG	153,42 8	116,12 9	1.374	1.374	+1
인하 /NNG	50,452	86,461	0.606	1.647	-1
...

2.3 Polarity Classification

Step 2. Naïve Bayes Classifier 모델링

1 < Intensity < 1.3 인 gray area 값들은 제외



2.3 Polarity Classification

Step 3. Dictionary

Hawkish	Dovish
국고채/NNG;금리/NNG;뛰/VV 지수/NNG;고용/NNG;지수/NNG;상승/NNG 오래/MAG;금리/NNG;인상/NNG 안정/NNG;심리/NNG;회복/NNG 실현지수/NNG 든게/MV 시끌벅적/MAG 높/VA;국제/NNG;유가/NNG;상승/NNG 높/VA;주가/NNG;오르/VV 콜/NNG;금리/NNG;인상/NNG;통화정책/NNG;불확실성/NNG 통화정책/NNG;불확실성/NNG;부담/NNG 국제채/NNG 인상/NNG;가능성/NNG;낮/VA;금리/NNG;인상/NNG 경제/NNG;성장률/NNG;호전/NNG 주택가격/NNG;내림/NNG 서비스업/NNG;지표/NNG;부진/NNG 판매/NNG;재고/NNG;증가/NNG 미/NNG;국채/NNG;가격/NNG;떨어/VV 과잉/NNG;유동성/NNG;경기/NNG;회복/NNG 불안요인/NNG;증가/NNG 주택/NNG;공급/NNG;과잉/NNG;우려/NNG	ecb/NNG;금리/NNG;인상/NNG;금리/NNG;인상 /NNG 수요/NNG;기대/NNG;어렵/VA 수요/NNG; 한계/NNG 안정/NNG;기대/NNG;확산/NNG 국채선물시장/NNG;인플레이션/NNG;우려/NNG 인플레이션/NNG;유가/NNG;하락/NNG 우려/NNG;심리/NNG;위축/NNG 전미소매업협회/NNG 비지표채권/NNG 채권시장/NNG;단기/NNG;금리/NNG;하락/NNG 봇물/NNG;터지/VV;쏟/VV 달러엔/NNG;환율/NNG;하락/NNG;압력/NNG 주택/NNG;판매/NNG;예상/NNG;하회/NNG 수출/NNG;출하/NNG;증가/NNG 총자본투자효율/NNG 하락/NNG;불구/NNG;높/VA 고용증가율/NNG; 둔화/NNG ...

2.3 Polarity Classification

Step 4. Measuring Sentiments

n-gram의 극성점수(polarity score)를 이용해 문장(sentences)와 문서(documents)의 톤 측정

i . 개별 문장(sentence)의 어조($tone_s$) 측정

$$tone_s = \frac{\text{No.of hawkish features} - \text{No.of dovish features}}{\text{No.of hawkish features} + \text{No.of dovish features}}$$

ii. 개별 문서(document)의 어조($tone_i$) 측정

$$tone_i = \frac{\text{No.of hawkishtone}_{s,i} - \text{No.of dovish tone}_{s,i}}{\text{No.of hawkish tone}_{s,i} + \text{No.of dovish tone}_{s,i}}$$

2.3 Polarity Classification

Step 4. Measuring Sentiments

n-gram의 극성점수(polarity score)를 이용해 문장(sentences)와 문서(documents)의 톤 측정

No.	Date	BOK minutes (Foreign Currency + Financial Markets)	$tone_i$
1	20050609	일부 위원은 우리나라 경제가 일본경제에 비해 더 나아질 특별한 요인이 없음에도 ...	0.172414
2	20050707	일부 위원은 6월 들어 국제유가가 크게 오르고 원화와 엔화간 동조화 현상이 뚜렷...	0.142857
3	20050811	일부 위원은 위안화 절상 이후 아시아 통화중 우리나라 원화의 절상폭이 가장 컸다...	0.380282
4	20050908	일부 위원은 최근 인도네시아 금융불안 사태가 발생한 가운데 국제 환투기세력이 인...	0.241379
5	20051011	일부 위원은 미국과 우리나라의 정책금리 격차가 확대되고 있는 가운데 금년 두 차...	0.473684
...

III. 프로젝트 결과



3. Project Outcome

프로젝트 결과

01

정확도(Accuracy) 측정

- Intensity에 따른 모델 정확도 측정

02

$MP\ Sentiments\ (tone_{mkt}) \leftrightarrow MP\ Decision, BOK\ Policy\ rate, EPU$

- BOK Policy rate / MP Decision
- EPU

03

$MP\ Sentiments\ (tone_{mkt}) \leftrightarrow Macroeconomic\ Variables$

- CPI
- Stock Market Price / 장단기 스프레드

3. Project Outcome

정확도(Accuracy) 측정

```
predict(test_set, testdf, testdf[“counting_h”].sum(), testdf[“counting_d”].sum(), 1.3)
```

Intensity = 1.3 | Accuracy = 69.800%

```
predict(test_set, testdf, testdf[“counting_h”].sum(), testdf[“counting_d”].sum(), 1.5)
```

Intensity = 1.5 | Accuracy = 71.396%

```
predict(test_set, testdf, testdf[“counting_h”].sum(), testdf[“counting_d”].sum(), 1.3)
```

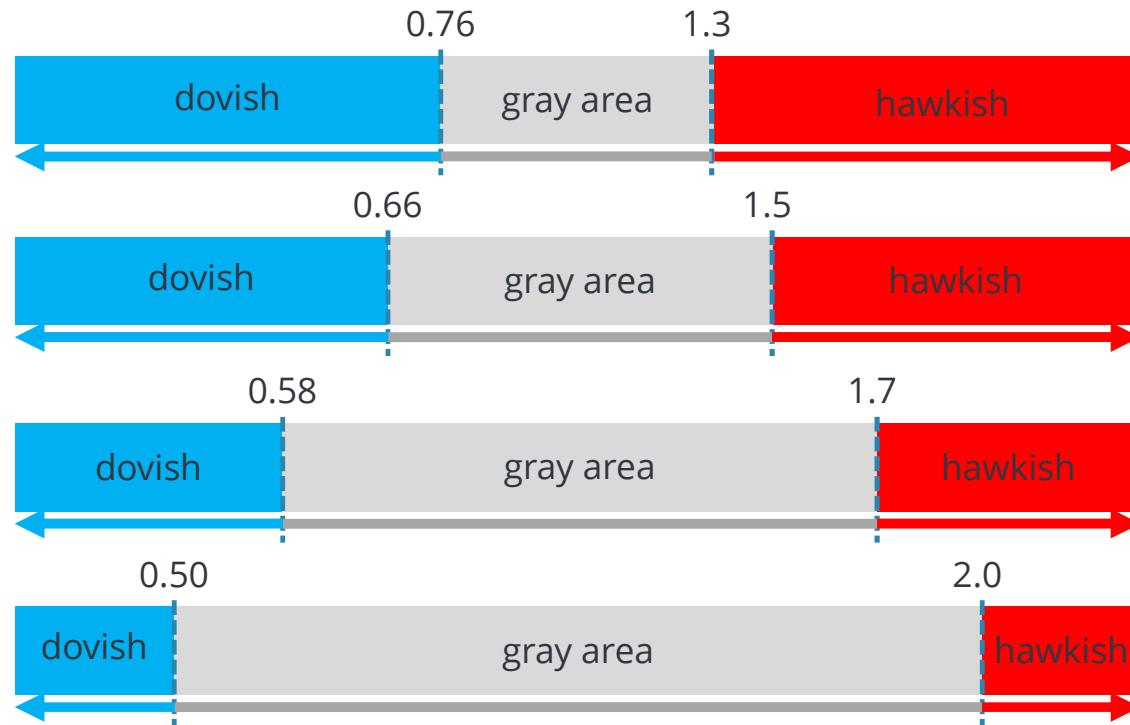
Intensity = 1.7 | Accuracy = 72.059%

```
predict(test_set, testdf, testdf[“counting_h”].sum(), testdf[“counting_d”].sum(), 1.3)
```

Intensity = 2.0 | Accuracy = 71.181%

3. Project Outcome

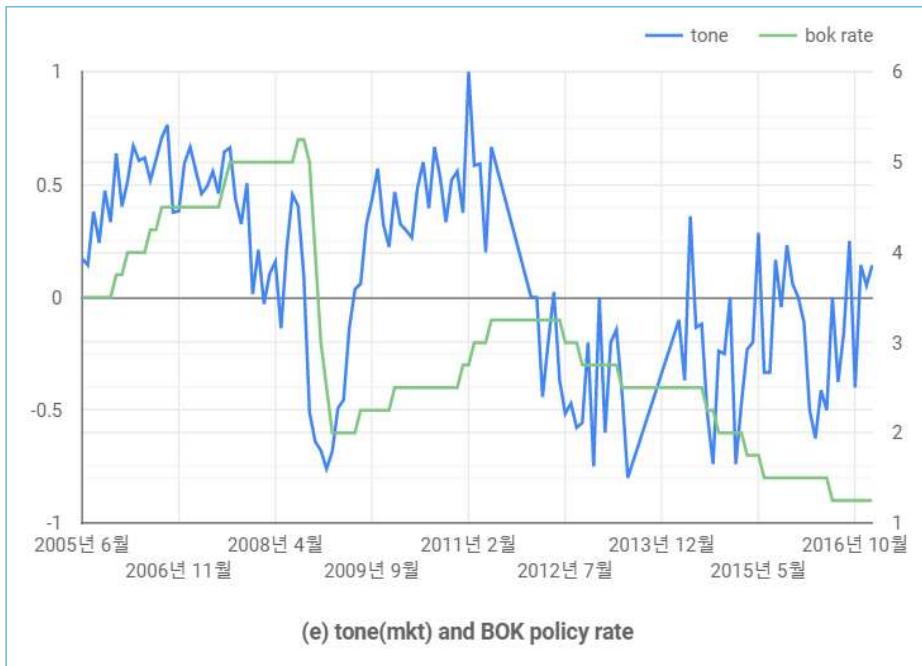
정확도(Accuracy) 측정



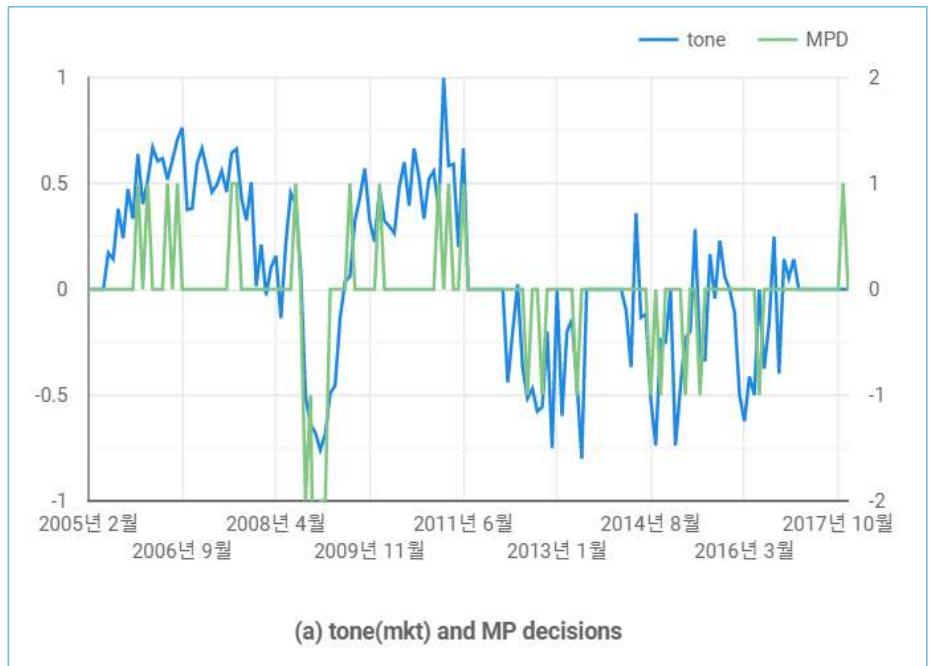
Intensity가 커질수록 Overfitting 발생 가능성 ↑

3. Project Outcome

프로젝트 결과 비교

(e) $tone_{mkt}$ and BOK policy rate

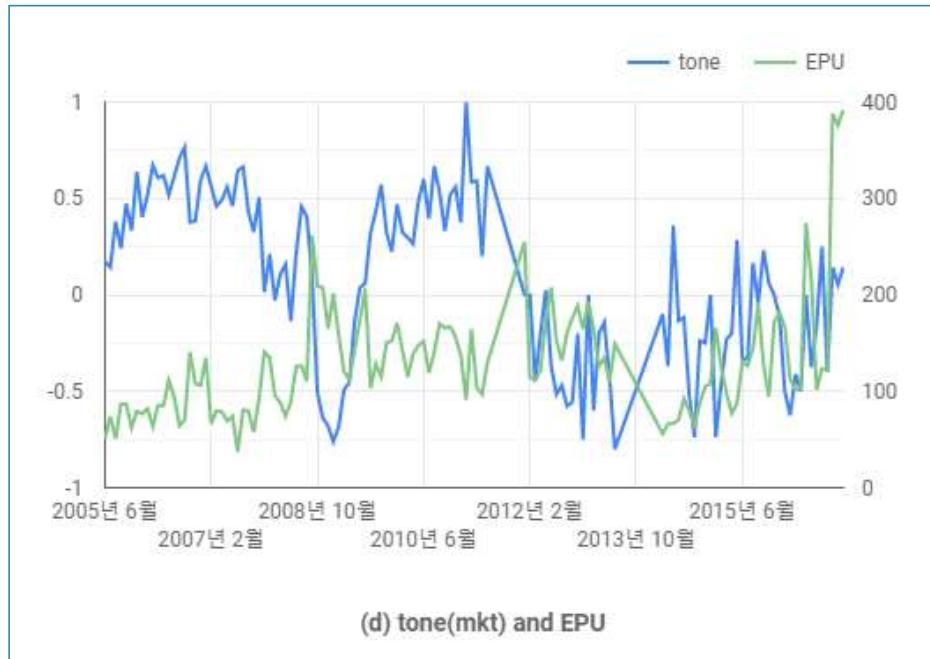
출처 : 한국은행

(a) $tone_{mkt}$ and MP decision

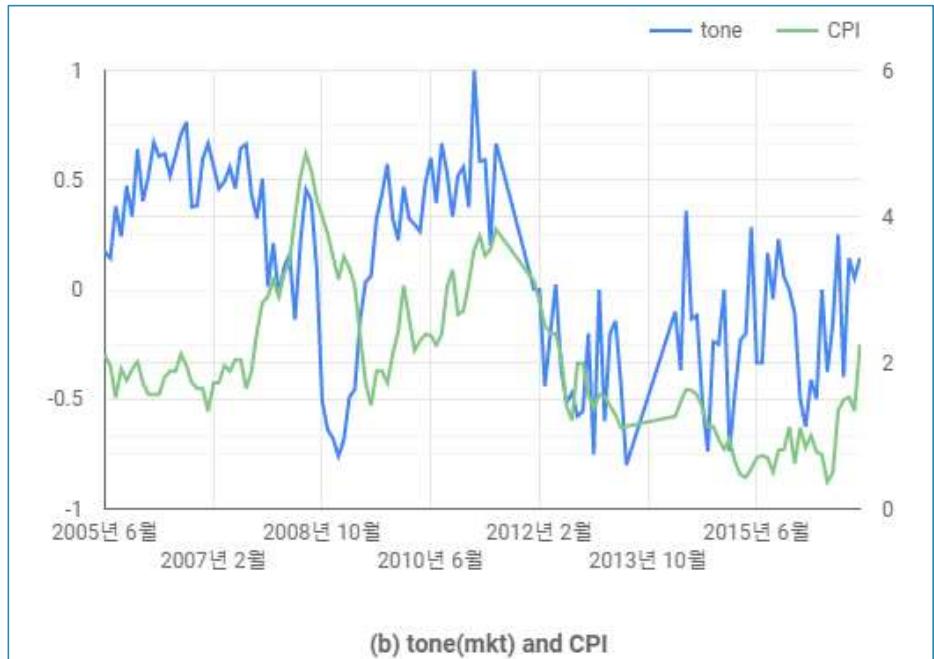
출처 : 한국은행

3. Project Outcome

프로젝트 결과 비교

(d) $tone_{mkt}$ and EPU

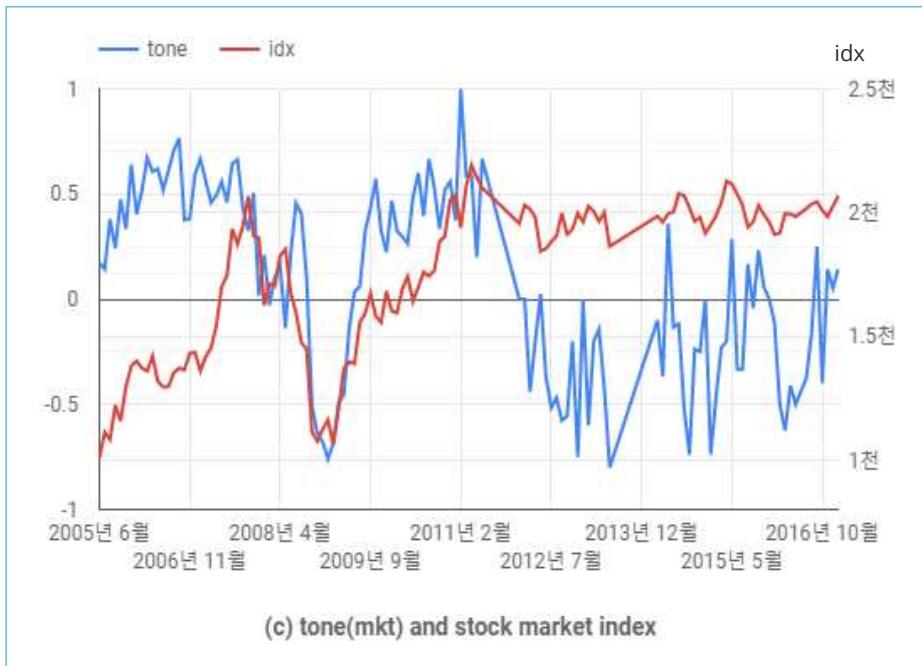
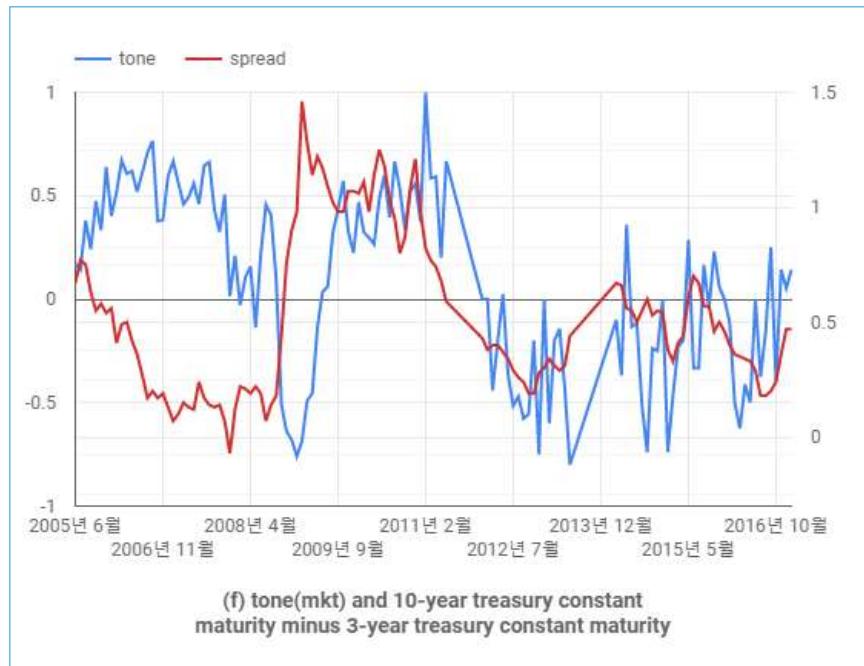
출처 :
http://www.policyuncertainty.com/korea_monthly.html

(b) $tone_{mkt}$ and CPI

출처 : 통계청

3. Project Outcome

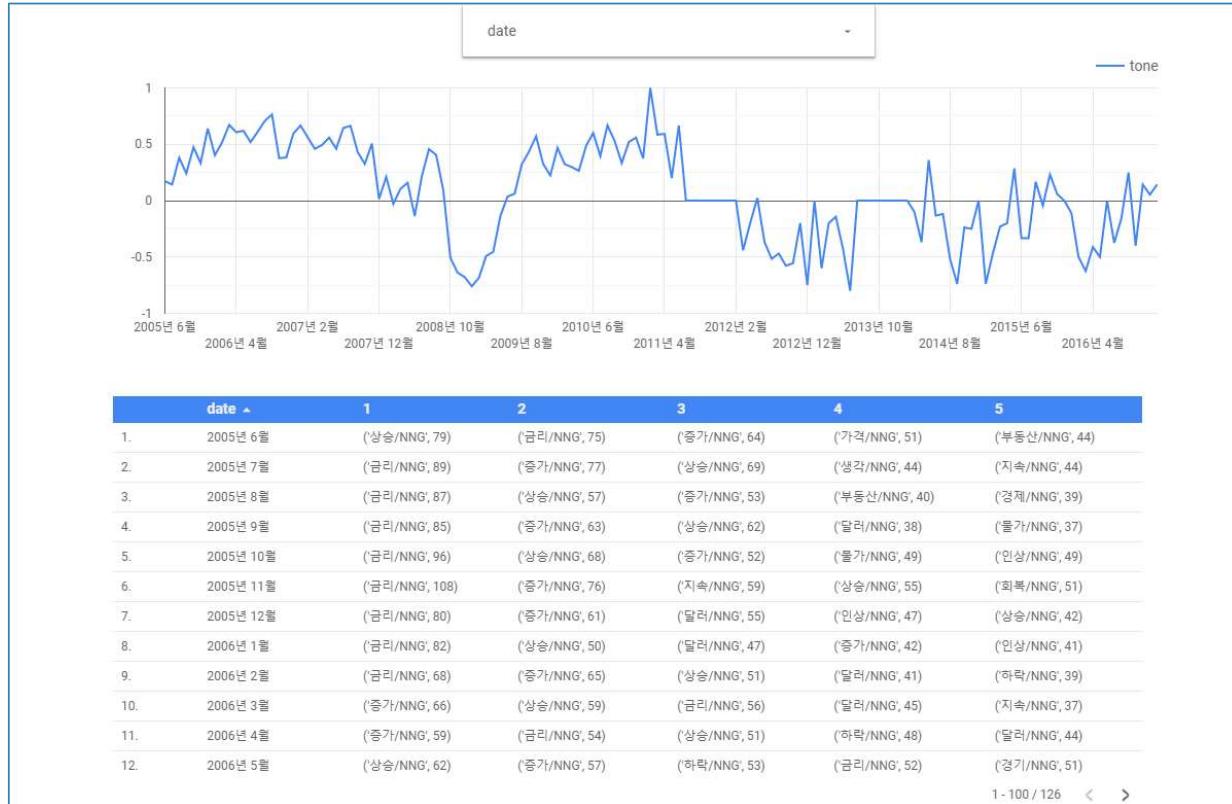
프로젝트 결과 비교

(c) $tone_{mkt}$ and stock market index출처 : <http://www.krx.co.kr/main/main.jsp>(f) $tone_{mkt}$ and 장단기 스프레드

출처 : e-나라지표

3. Project Outcome

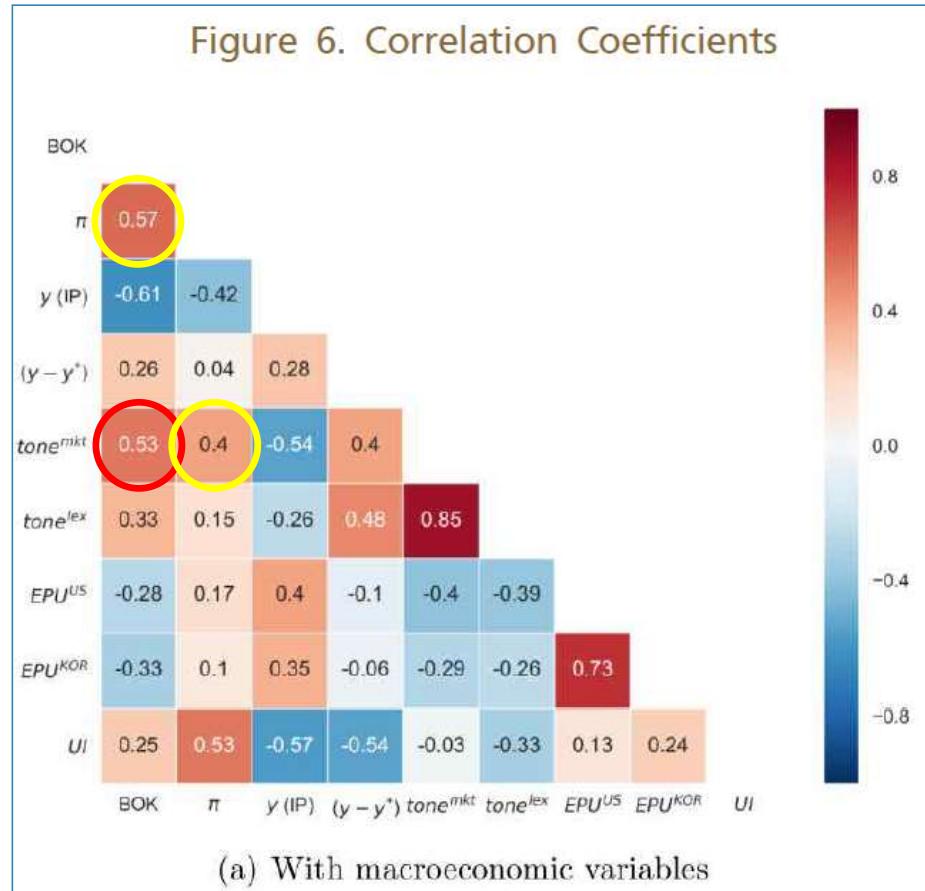
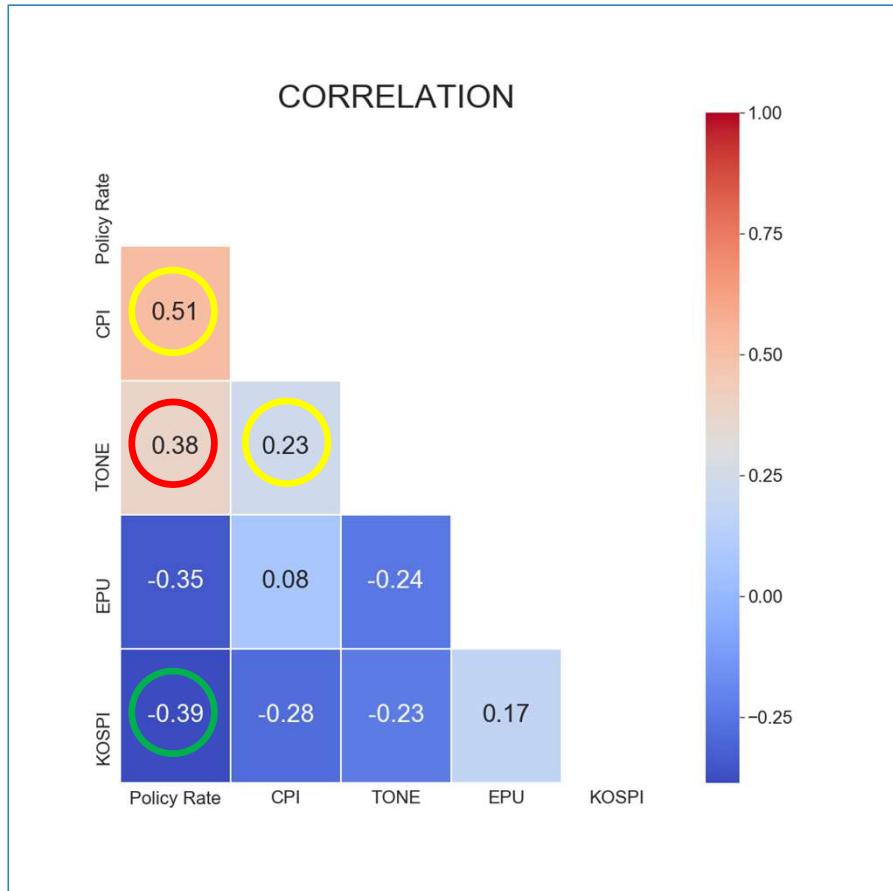
프로젝트 결과 비교



데이터스튜디오 : shorturl.at/ovCQ3

3. Project Outcome

Correlation Coefficients



3. Project Outcome

Wordclouds



IV. 프로젝트 시사점



4. Project Implications

< 프로젝트 목적 >



1. 데이터 처리부터 분석까지 논문을 충실히 따르며 직접 구현
2. 프로젝트 결과물의 활용 및 응용 가능성 검토

4. Project Implications

논문에서 언급한 연구의 시사점

텍스트 마이닝을 이용한 금통위 의사록 분석

- ① 통화 정책 효과성 분석 모델에 적용 가능
- ② 장래의 통화 정책 결정 및 한은의 의사 소통의 효과성 평가
- ③ 거시경제 불확실성, 미래의 통화 정책 기조에 대한 대중의 기대, 주식 시장 정서 등을 측정에 활용
→ 자산가격이나 실제 변수에 미치는 영향 조사

4. Project Implications

프로젝트 후속 연구 방향

텍스트 마이닝을 이용한 기준금리 예측

- ① 통화 정책 효과성 분석 모델에 적용 가능
- ② 장래의 통화 정책 결정 및 한은의 의사 소통의 효과성 평가
- ③ 거시경제 불확실성, 미래의 통화 정책 기조에 대한 대중의 기대, 주식 시장 정서 등을 측정에 활용
→ 자산가격이나 실제 변수에 미치는 영향 조사

4. Project Implications

프로젝트 후속 연구 방향

- ③ 거시경제 불확실성, 미래의 통화 정책 기조에 대한 대중의 기대, 주식 시장 정서 등을 측정에 활용
→ 자산가격이나 실제 변수에 어떻게 영향을 미치는지 조사 가능

주식시장

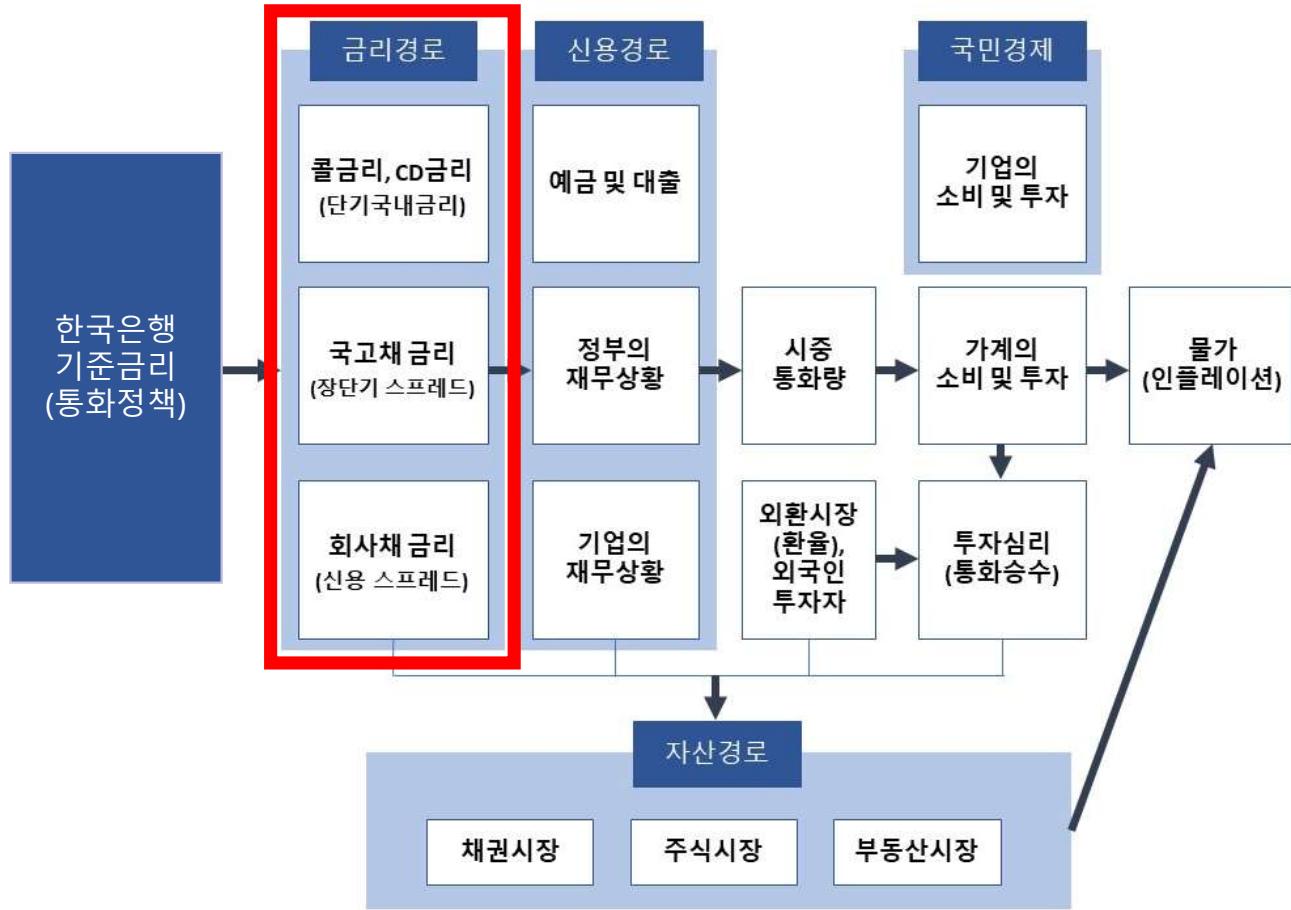
- 소규모 개방 경제인 국내 주식시장에 부적합
- 기준금리 이외에 수많은 변수가 존재

채권시장

- 기준금리 인상 시 직접적인 영향을 받음
- 논문 근거로 한 예측의 의미가 높음
- 채권투자, 금리나 채권을 기초자산으로 하는 파생상품 투자에 활용

4. Project Implications

프로젝트 후속 연구 방향



4. Project Implications

프로젝트 후속 연구 방향

매일경제

2019년 06월 01일

시장 즉각반응...치솟은 채권값

"새로운 인하 사이클"
20년·30년물 금리도
기준금리 아래로 '뚝'

한국은행 금융통화위원회에서 기준금리 인하 소수 의견이 나왔다는 소식에 채권금리가 요동쳤다. 그동안 시장 일각에서 연내 기준금리 인하 가능성을 꾸준히 제기하긴 했지만 이번 금통위에서 소수 의견이 등장하자 금리 인하 전망이 더욱 힘을 얻는 분위기다. 시장에서는 하반기 기준금리가 하향 조정된다면 새로운 금리 인하 사이클의 시작으로 봐야 한다는 의견도 나온다.

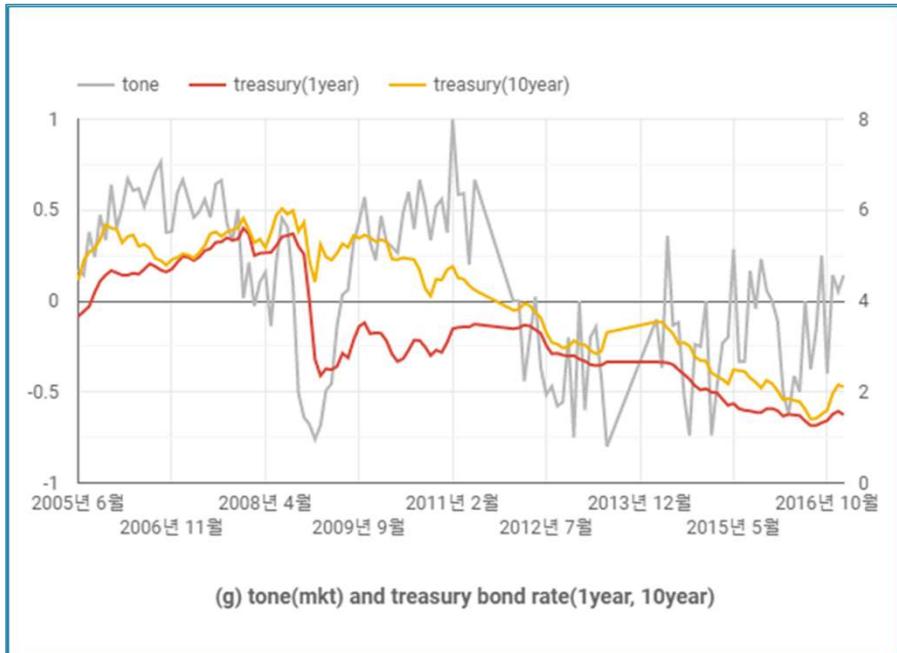
10년물 금리 역시 전 거래일 대비 5.9bp 떨어진 1.682%로 마감했으며, 3년물 금리는 1.587%로 거래를 마쳐 1.6%를 하회했다. 이날 **이주열 한은 총재는 소수 의견과 금통위 의견은 다르다며 '선 긋기'에 나섰지만 채권시장에는 소수 의견 영향이 더욱 강하게 작용했다.**

문홍철 DB금융투자 연구원은 "기존에도 채권시장에는 기준금리 인하 기대감이 녹아 있었는데, 소수 의견으로 더욱 강해진 것으로 보인다"며 "미·중 갈등이 장기화하며 불확실성이 더욱 커지는 상황에서 보험성 성격을 지닌 채권이 더욱 강세를 보이고 있다"고 설명했다.

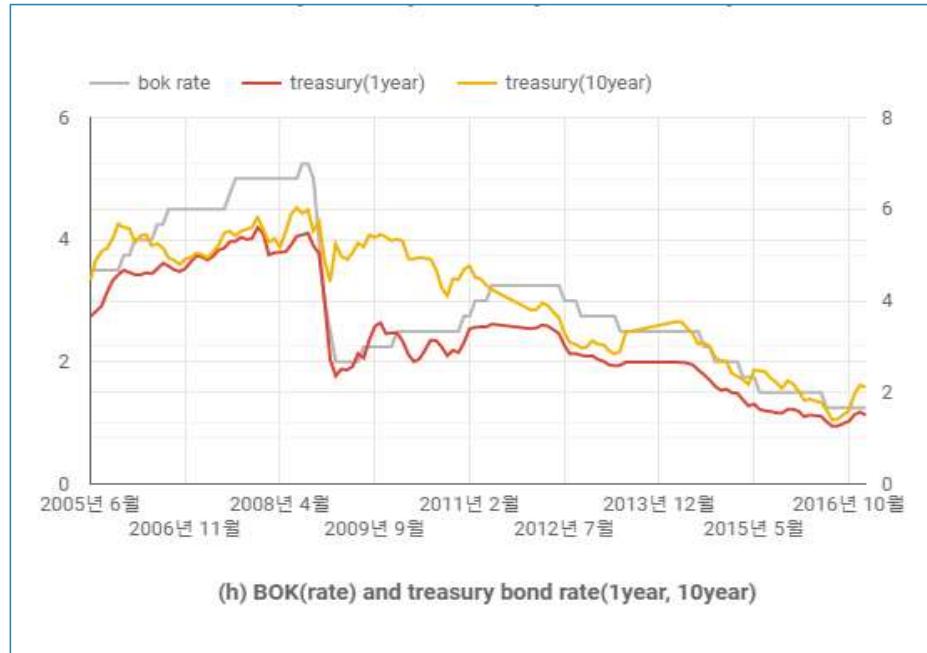
4. Project Implications

프로젝트 후속 연구 방향

(a) tone(mkt) and treasury bond rate(1year, 10year)



(b) BOK policy rate and treasury bond rate(1year, 10year)



V. 프로젝트 정리



4 프로젝트 정리

목 표

- Market Approach를 중심으로 논문을 충실히 따르며 직접 구현
- 비정형 텍스트 분석을 코드로 구현

잘한 점

- 목표 설정을 통해 짧은 기간 프로젝트 완수
- 적절한 업무분담으로 진행속도 향상

부족한 점

- Naïve Bayes Classifier 모델의 정확도 차이(69.8% ↔ 86%)
- 채권 분석 보고서 문서 개수 부족(3,095 ↔ 25,325)
- 시각화 및 포트폴리오 자료용 데이터 재처리로 일정 지체
- '왜 경제 신문사 기사가 포함되지 않았는가'라는 의문을 해소하지 못함

완료	to-do	
지체	.이데일리, 연합인포 텍스트 전처리 코드	
행색	.연합뉴스 텍스트 전처리 코드	
영역	.채권보고서 텍스트 전처리	
진아	.이데일리, 연합인포 텍스트 전처리 코드	
하영	.의사록 텍스트 전처리 코드	
다같이		
no	TO DO	데이터(code, csv 결과물) 위치 정리 완료된 코드 위치:
1	뉴스기사 csv 파일로 변환	52.141.32.158:8888/def/3팀/codes_final COMPLETE 1.crawling-edaily.ipynb
2	MPB 의사록 section 2.3만 줍기	COMPLETE 0_금통위의사록수집.ipynb : def function
3	뉴스 기사 정제 작업	/COMPLETE regexp_jh190521_yeonhap_info_reexp.ipynb
4	채권보고서 정제 작업	csv : 3팀/DATA/NBC/bond_pre.csv
5	글리 데이터 수집(플러그리, 기준글리)	csv : 3팀/DATA/rate.csv
6	corpus 정제(2005-2017) 데이터 확인	csv : 3팀/DATA/NBC/NBC_final/news_processed(2005-2017)_step1.csv
7		
8	Pre-processing	
9	tokenizing	
10	pos 품사태깅	
11	lemmatization	
12	불용어 처리	
13	word list 작성	
14	n-gram화 (5-gram, 15번 미만 나온 n-gram 삭제)	COMPLETE 2&3_token_ngram.ipynb csv : 3팀/DATA/NBC-merged.csv
15	market approach (마신러닝)	
16	줄거리 데이터 준비(동락문에 따른 반별정)	COMPLETE 4.0.1_mktapproach_labeling-ngramunigram.ipynb csv : 3팀/DATA/NBC/NBC_step1.ipynb
17	step1. tokens + n-gram 결합 합치기	COMPLETE 4.2_mktapproach_counting.ipynb csv : 3팀/DATA/NBC_step2-190529.csv
	step2. counting / 조건부확률 계산	COMPLETE NBC_step3(making dictionary).ipynb 서비스/大局(DATA/NBC_step3(dictionary).final-190529.csv)
	step3. 매파/비율기파 dictionary	COMPLETE 4.4_mkt_approach_tone.ipynb csv : 3팀/DATA/tone-simple-190529(final).csv

<프로젝트 단계별 목표 설정 스프레드 시트>

4 프로젝트 정리

분석 도구

- 언어 : Python
- 개발환경 : Jupyter Notebook
- 라이브러리 : Pandas, Numpy, Sklearn, mecab_eKoNLPy,

결과물

- Git Hub : <https://github.com/nlpbokproject/jh>
- Google Data Studio :

참고 논문

- Deciphering Monetary Policy Board Minutes through Text Mining Approach
- Ngram2vec- Learning Improved Word Representations from Ngram Co-occurrence Statistics
- Inducing Domain-Specific Sentiment Lexicons from Unlabeled Corpora

참고사이트

- 한국은행 <http://www.bok.or.kr/>
- e-나라지표 <http://www.index.go.kr/main.do?cate=6>
- 픽사베이 <https://pixabay.com/ko/>
- 한국자금증개 <http://www.kmbco.com/bond/>
- 한국거래소 <http://www.krx.co.kr/>
- Economic Policy Uncertainty <http://www.policyuncertainty.com/>



THANK YOU