

텍스트 마이닝을 활용한 시장지표 예측 모델링

Index

1. 관련 논문

2. 절차

- 데이터 수집
- 데이터 클렌징
- 데이터 라벨링
- 통계 & 딥러닝 모델링

3. 통계기반

- 전처리
- Feature Selection
- NBC
- Predict

4. 딥러닝기반

- 전처리
- Modeling
- Train
- Predict

Article

관련 논문

텍스트 마이닝을 활용한 한국은행 기준금리 예측

- ['Deciphering Monetary Policy Board Minutes through Text Mining Approach: The Case of Korea'](#)

- 요약본

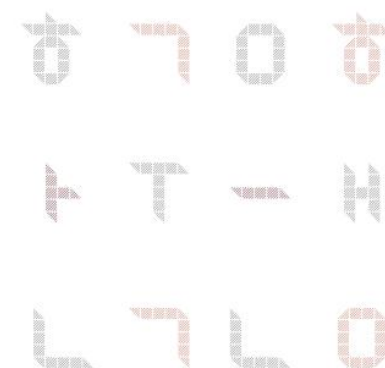
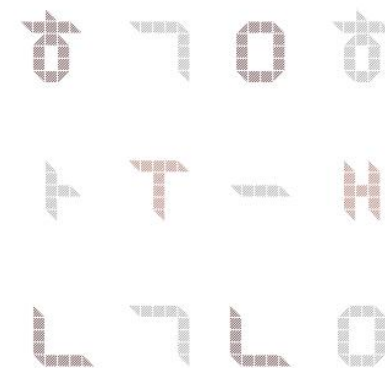
No. 2019-1

BOK Working Paper

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

Ki Young Park, Youngjoon Lee, Soohyon Kim

2019. 1



Process

절차

데이터 수집

- [기준금리](#)
- [콜금리](#)

- [한국은행 의사록](#)
- [채권보고서](#)
- [뉴스기사](#)

- 권장 수집량 : 뉴스기사 20만 건 이상
- 기술 스택 : Scrapy, 멀티스레딩

금융통화위원회 의사록

🏠 > 한국은행 > 금융통화위원회 > 금융통화위원회 의사록

금융통화위원회 의사록

Total : 495건 [1/50 pages]

🔍 검색

📄 상세검색

금융통화위원회 의사록(2022년도 제19차)(2022.10.12)

🕒 2022.11.01

👁 1868

금융통화위원회 의사록(2022년도 제18차)(2022.9.22)

금융홈 > 리서치 > 채권분석 리포트

채권분석 리포트

채권분석 채권대학살 이후

2023년 미국 국채시장 전망2022년 긴축 효과가 인플레이션 둔화로 연결되는지 주목해야겠다. 고물가 & 긴축 & 이연소비 약화가 기업 실적 악화, 실업을 상승을 거쳐 수요축 인플레이션 완화로 연결될 수 있다. 내년 물가 하락세와 함께 경기 하강 흐름도 동반될 것이다. 신한투자증권 | 2022.11.04

제목	증권사	첨부	작성일	조회수
주간 채권 코멘트 (11월 4일) 📄	신한투자증권	📄	22.11.04	139
채권 Daily (11.04) 📄	유안타증권	📄	22.11.04	215
채권대학살 이후 📄	신한투자증권	📄	22.11.04	337
유틸(U-turn) 신호 기다리기 📄	신한투자증권	📄	22.11.04	106
채권 Daily (11.03)	유안타증권	📄	22.11.03	449

PICK 언론사가 선정한 주요기사 혹은 심층기획 기사입니다.

📰 연합뉴스 PICK 1일 전 네이버뉴스

美 연준, 4연속 '자이언트 �텝'...한은도 24일 금리인상 확실시(중...

美 금리 3.75~4.00%로 올라 15년만에 최고...韓보다 최대 1.0%p 더 높아 파월, 이르면 12월 속도조절 언급..."최종금리 더 높아질 것" 5% 육박 시사 이상헌 특파원 ...

미 연준 4차례 연속 '자이언트 �텝'...기준... 한겨레 PICK 1일 전 네이버뉴스
미 연준, 4연속 '자이언트 �텝'...한국은행도... JTBC PICK 1일 전 네이버뉴스
연준, 4연속 '자이언트스텝'...美 기준금리 ... 더팩트 PICK 1일 전 네이버뉴스
파월 "더 높게, 더 오래"...최종금리 5... 서울경제 PICK 23시간 전 네이버뉴스

관련뉴스 6건 전체보기 >

📰 한국경제 PICK 20시간 전 네이버뉴스

영국도 '자이언트스텝'...33년 만에 금리 최대폭 인상

영국 중앙은행이 기준금리를 0.75%포인트 올리는 '자이언트스텝'을 단행했다. 전날 미국 중앙은행(Fed)의 기준금리 인상에 보조를 맞춘 것으로 풀이된다. 영국 중앙...

영국중앙은행, 기준금리 0.75%p 인... 조선비즈 PICK 20시간 전 네이버뉴스
英 중앙은행, 기준금리 3%로 0.75... 아시아경제 PICK 20시간 전 네이버뉴스
영 중앙은행도 '자이언트 �텝'...기... 연합뉴스 PICK 20시간 전 네이버뉴스
영국도 금리 '자이언트 �텝'...장기... 연합뉴스 PICK 17시간 전 네이버뉴스

관련뉴스 8건 전체보기 >

데이터 클렌징

- 수집된 데이터에서 분석에 필요 없는 부분을 제거하는 과정
- 수집된 데이터를 눈으로 꼭 확인하고 클렌징 코드를 작성할 것
(뉴스 언론사별로 다양한 형태의 불필요한 내용이 있을 수 있음)
- 정규표현식 혹은, replace를 활용



뉴욕증권거래소의 트레이더

[AFP 연합뉴스 자료사진. 재판매 및 DB 금지]

삭제

(서울=연합뉴스) 임상수 기자 = 최근 금융시장 혼란 이후 시장에서는 미 연방준비제도(Fed·연준)가 공격적인 금리인하로 경기침체 우려를 잠재우기를 원하고 있다.

데이터 라벨링

- 학습데이터를 구성하는 단계로 목적에 따른 label 전략을 수립
- 목적 : 하루 뒤의 금리가 어떻게 변하는지, 한 달 뒤의 금리가 어떻게 변하는지, 반 년 뒤의 금리가 어떻게 변하는지

기준 금리 변동 분석 (Sentiment Analysis)

- 통계 기반 : 시장 접근법 (NBC)
- 사전 접근법 (SO-PMI)
- 딥러닝 접근법 : 작은 딥러닝 자연어처리 모델 활용
 - KoBERT, KoELECTRA 등...
- LLM AI : Gemini, GPT API를 활용

Statistical

통계기반 모델링

전처리

- 문장 분리
- 통계기반 텍스트 전처리
- eKoNLPy

Feature Selection

- N-gram

성능 최적화

- 로직 직접 구현
- 파이썬 내장함수 활용
- 멀티프로세싱 활용

모델링

- NBC

예측 & 평가

- 기준금리의 상승, 하락을 맞췄는지
- 앞으로 기준금리가 어떻게 될지 예상

AI

LLM 모델링

프롬프트 작성

- AI가 효과적으로 문서를 분석하여 도출할 수 있도록 분석방향 및 기준, 제약조건, Output format을 정리하여 프롬프트를 작성

Parsing

- Json 포맷, 텍스트 포맷, CSV포맷 등 LLM이 출력으로 반환한 정보를 파이썬 객체로 변환하여 활용할 수 있도록 Parsing

예측 & 평가

- 기준금리의 상승, 하락을 맞췄는지
- 앞으로 기준금리가 어떻게 될지 예상

산출물

- Notion (프로젝트 기획, 연구내용 정리)
- Github (코드 관리)
- Google Drive (데이터, 파일 관리)

- 위에서 작성한 내용을 토대로 PPT에 정리