수집된 데이터 및 데이터 전처리 문서

SKN Family AI Camp 2기: 최종 프로젝트 2조



프로젝트 주제

경제 지식이 없는 국민들을 대상으로 경제 지식을 쉽게 설명하는 AI 인플루언서 웹 페이지 관리 Tool

[한국은행 제공 경제교육용 PDF 문서]

- <u>경제금융용어 700선</u>
- <u>알기 쉬운 경제 이야기 (일반인용)</u> 수집 데이터

[네이버 경제 뉴스: 최신 이슈 Top 3]

-<u>SBS 경제 뉴스</u>

[원본]



경제금융용어 700선

[전처리 결과 데이터 : CSV 파일]

	title	content	related_keyword
0	가계부실위험지수(HDRI)	가구의 소득 흐름은 물론 금융 및 실물 자산까지 종합적으로 고려하여 가계부채의 부실	총부채원리금상환비율(DSR)
1	가계수지	가정에서 일정 기간의 수입(명목소득)과 지출을 비교해서 남았는지 모자랐는지를 표시한	경상수지, 재정수지
2	가계순저축률	일반적으로 저축률은 저축액을 처분가능소득으로 나눈 비율을 말한다. 마찬가지로 가계순	NaN
3	가계신용통계	가계신용통계는 가계부문에 대한 신용공급 규모를 나타내는 통계이다. 가계신용은 금융기	NaN
4	가계처분가능소득	가계처분가능소득(PDI; Personal Disposable Income)은 가계가	국민총소득(GNI)
695	TED 스프레드	TED 스프레드는 원래 미국 단기국채(T-Bill) 선물금리와 유로달러 선물금리 간	LIBOR, Treasury Bill(T/B)
696	Treasury Bill(T/B)	Treasury Bill(T/B)은 미국 재무성이 발행하는 만기 4주(1개월), 1	NaN
697	VAN사업자	통신사업자로부터 회선을 대여 받아 고도의 통신처리기능으로 부가가치를 높여 서비 스를	지급결제시스템, 전자금융
698	VaR(Value at Risk)	주어진 신뢰수준 하에서 일정 기간 동안 발생할 수 있는 '최대 손실금액'으로 금융기	예상손실
699	VIX	미국 주식시장의 단기 변동성에 대한 시장의 기대치를 나타내는 지수로 시카고옵션거 래	NaN

700 rows X 3 columns

[전처리 결과 데이터 사용 목적]

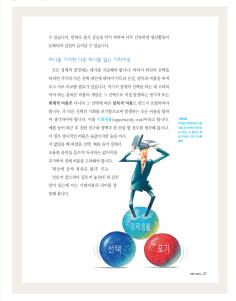
위 데이터는 '오늘의 단어' 카테고리에 쓰일 예정입니다.

'오늘의 단어'는 한국은행에서 발행한 '경제금융용어700선'에 실린 단어들을 쉽게 설명해 주는 일일 컨텐츠입니다.

기존 설명은 경제 지식을 잘 모르는 일반인이 보기에 무리가 있습니다.

따라서 LLM의 fine-tuning 기능, Auto-RAG 등의 기법을 통해 쉬운 설명으로 변환시킨 후, 일반인들에게 소개하고자 합니다.

[원본]



알기 쉬운 경제 이야기 (일반인용)

[전처리 결과 데이터 : JSON 파일]

객체 개수 : 139개, key 개수 : 3개 (subject, title, content)

[전처리 결과 데이터 사용 목적]

위 데이터는 메인 컨텐츠를 위해 사용할 예정입니다.

메인 컨텐츠는 'LLM 인플루언서'의 친근함을 매개로 '알기 쉬운 경제 이야기(일반인용)'을 활용 및 재편성할 예정입니다.

알기 쉬운 경제 이야기는 여러 내용들을 엮어야 하기 때문에, Parsing 작업에 유리한 JSON 파일 형식으로 선택했습니다.

('경제금융용어 700선' 활용 컨텐츠는 여러 내용을 엮을 필요 없이 row 1개만 써도 무방해서 CSV 파일 형식을 채택했습니다.)

메인 컨텐츠는 LLM의 fine-tuning 기능, Auto-RAG 등의 기법을 통해 문장들을 새롭게 엮음으로써 다양한 컨텐츠를 보여주고자 합니다.

[원본]



SBS 경제 뉴스

[전처리 결과 : CSV 파일]

	origin_title	origin_body	link	date
0	삼성전자, 포브스 선정 '세계 최고 직장'서 3위로 하락	삼성전자가 3분기 실적 부진 등으로 위기감이 고조된 가운데 미국 경제지 포브스가 선	https://n.news.naver.com/article/055/000119614	20241010
1	외국인 매도 행렬1년 7개월 만에 '5만전자'	<앵커>\n\n이를 전 예상을 밀도는 3분기 실적을 내놨던 삼성전자의 주가가 1년	https://n.news.naver.com/article/055/000119636	20241010
2	[친절한 경제] 이자 줄고 해외직구 늘고?WGBI 편입 의미와 효과	<앵커>\n\n목요일 친절한 경제 권애리 기자와 함께합니다. 권 기자, 우리나라가	https://n.news.naver.com/article/055/000119612	20241010

[전처리 결과 데이터 사용 목적]

위 데이터는 경제 트렌드를 반영하기 위해 수집한 SBS 경제 기사 크롤링 데이터입니다.

경제 섹션에서 하루 동안 이슈가 많이 되었던 기사 Top 3를 수집합니다.

수집한 기사들은 LLM의 fine-tuning 기능과 Prompt-Engineering을 통해 일반인들이 이해하기 쉬운 어휘로 가공하여 보여주고자 합니다.