

프로젝트 기획 및 RAG 파이프라인

서비스명 및 개요

서비스명: 나의 투자 친구, 머니또

표어: 나에게 돈이 또?

서비스 개요:

- 서비스가 해결하려는 문제점(**Problem Statement**)
 - 2030 투자를 처음 시작하는 투린이에게 부족한 경제 지식, 투자 상품에 관한 정보 제공 및 실시간 이슈를 포함한 투자 종목 추천
 - 최근 금융, 정치, 전쟁, 가상자산 등의 다양한 이슈로 인해 많은 청년들이 주수입원 외에도 금융상품에 투자해 추가적인 수익을 창출하려는 움직임을 보이고 있습니다. 그러나 2030세대 초보 투자자(투린이)는 경제 지식과 투자 상품에 대한 정보가 부족하여 시작 단계에서 어려움을 겪고 있습니다. 전문가 상담은 비용과 접근성 문제로 진입장벽이 높으며, 관련 이슈를 통합적으로 확인할 수 있는 플랫폼의 부재를 확인하였습니다.
- 제공하고자 하는 주요 기능 및 가치

'나의 투자 친구, 머니또' 서비스는 최근 해당 금융상품과 관련 있는 이슈를 검색하여, **Open API**의 챗지피티 서비스를 이용하여 신뢰도 높은 답변을 생성하는 챗봇 서비스를 통해 투자 도움을 제공합니다.

 1. 뉴스 기반 투자 정보 제공
: 최신 금융, 기업, 정치 이슈와 관련된 뉴스를 검색 및 요약하여 사용자에게 신뢰도 높은 투자 관련 답변 제공
 2. 투자 상품 & 종목 설명 및 추천
: 초보 투자자를 위한 경제 및 금융 기본 지식 제공
 3. 사용자 맞춤 피드백
: 사용자의 투자 성향 분석 및 목표에 부합하는 투자 전략과 정보를 정리해주는 개인화 서비스
- 타겟 사용자와 예상 사용 시나리오
 1. 타겟 사용자
 - 20~30대 초보 투자자(투린이)
 - 경제 및 금융 지식이 부족하지만 투자에 관심이 있는 사용자
 - 투자 정보를 빠르게 얻고 싶어 하는 직장인 및 대학생
 2. 예상 사용 시나리오
 - 사용자 A (직장인):
최근 뉴스에서 "전쟁 리스크"가 투자 시장에 미치는 영향을 확인하고 싶음.
챗봇에서 관련 뉴스를 요약받고, 추천 종목과 투자 전략에 대한 답변을 얻음.

	<ul style="list-style-type: none">- 사용자 B (대학생): 주식 투자를 시작하려는 대학생이 "삼성 관련 최신 뉴스"를 챗봇에 질문. 챗봇이 요약된 뉴스와 기본 주식 투자 가이드를 제공.- 사용자 C (초보 투자자): 현재 보유 중인 금융상품과 관련된 최신 뉴스를 검색. 관련 정보를 기반으로 추가 투자 상품 추천을 받음.
목표 및 기대효과	<p>서비스 목표:</p> <ul style="list-style-type: none">- 초보 투자자들이 자신의 투자 성향을 파악하고, 개인화된 금융 상품 추천을 통해 안정적이고 자신감 있는 투자 결정을 내릴 수 있도록 돕는 것 <p>기대효과:</p> <ol style="list-style-type: none">1. 경제에 관심 없는 인구의 경제활동 유도<ol style="list-style-type: none">a. 경제나 투자를 몰랐던 사람들에게 친근하고 쉬운 접근법을 제공하여 자연스럽게 경제활동에 참여하도록 유도.b. 복잡한 금융 상품을 이해하기 쉽게 설명하여 금융문맹(Financial Illiteracy) 해소에 기여.2. 금융 시장의 다양성과 경쟁력 강화<ol style="list-style-type: none">a. 초보 투자자 유입 증가로 인해 금융 시장 활성화 및 신규 투자 상품 개발 촉진3. 초보자를 위한 안전한 금융 길잡이 역할<ol style="list-style-type: none">a. 실제 금융 상품에 기반한 자료로 불법적인 금융 상품이나 사기 피해를 감소b. 개인의 소비 패턴, 위험 성향, 목표 등을 분석해 적합한 금융 상품을 추천, 불필요한 시행착오를 줄임4. 최신 정보를 이용하여 해당 시점에 알맞는 투자 상품 탐색<ol style="list-style-type: none">a. 최신 경제 뉴스를 기반으로 단기적인 시장 변화 반영 및 경기 분석b. 실시간 경기 지수를 활용하여 최적 상품 제공
타겟 사용자 및 시장분석	<p>타겟 사용자: 2030 청년층, 투린이, 투자 경험 無</p> <ul style="list-style-type: none">- 예상 사용자 유형: 투자 경험이 전무한, 투자에 입문하고 싶은 대학생, 주부, 직장인- 주요 요구사항 및 사용 목적<ul style="list-style-type: none">● 사용자의 “투자 성향”과 “투자 니즈”를 파악합니다.● 실시간 뉴스 데이터를 활용하여 “맞춤형 투자 상품”을 추천합니다.● 투자 상품에 대한 설명을 제공하여 투자에 효과적으로 입문할 수 있도록 돕습니다.

데이터 구성 및 처리 방법

원천데이터 소스: 네이버 뉴스 기사, KDI 시사 경제 용어사전, 한국거래소(정보데이터시스템), Investing.com (<https://kr.investing.com/>), 전국투자자교육협의회 초보투자자 길라잡이, 토스 피드

원천데이터 형식: HTML, PDF 문서 파일

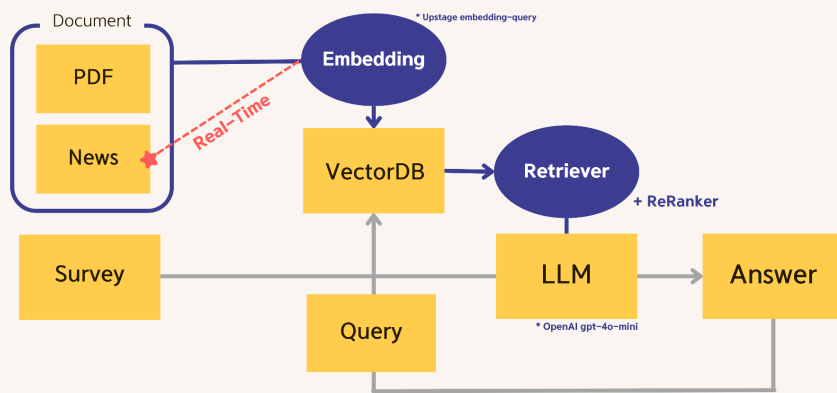
원천데이터 처리 방법:

- 데이터 수집
 - 검색 API를 활용한 뉴스 기사 수집
 - HTML → JSON 으로 변환하여 데이터베이스에 저장
 - 추출한 텍스트를 JSON 또는 데이터베이스 테이블로 변환
- 데이터 처리
 - PDF 내의 이미지를 텍스트로 변환 및 선별 텍스트 벡터화
 - 불필요한 HTML 태그, 특수 문자 제거 및 불용어 처리
 - 결측값 처리 및 데이터 클렌징

RAG

파이프라인 설계

작동 과정.



1. 데이터 최적화:
 - a. Text splitting
 - i. Chunk Size: 1000
 - ii. Overlap: 100
 - b. 질문 키워드 추출
 - i. 사용 도구 : Kiwi (한국어 키워드 추출)
2. 벡터 데이터베이스 구축 및 임베딩
 - 벡터 DB: Pinecone
 - 임베딩 모델: Upstage Solar Embedding
3. Retriever 및 Reranker 구현:
 - VectorStore Retriever 이용
 - 하이퍼파라미터 튜닝:
 - 반환할 문서 수(k) = 3
 - Reranker:
 - CrossEncoderReranker를 활용해 Retriever의 후보 문서를 재정렬 • 적용
 - HuggingFace의 'BAAI/bge-reranker-base' 모델 사용
 - 문서 수: 상위 3개 문서
4. LLM 프롬프트 설계 및 답변 생성, 평가:

- a. Task: QA / 챗봇
b. 프롬프트 예시 - 투자 성향에 따른 금융상품 추천

```
first_template = f"""
```

당신은 금융 상담 전문가이며, 고객의 투자성향에 맞는 투자상품을 참조문서에 기반해 추천하는 역할을 맡고 있습니다.

고객의 투자성향은 {customer_type}형입니다.

""" + """

1. 참조문서에서 고객의 투자성향을 찾아 추천되는 투자상품을 찾습니다.

2. 추천하는 투자상품 종류를 나열하고, 이후 투자상품과 그 특징을 설명합니다.

- 어려운 단어는 사용자가 이해하기 쉽도록 풀어서 설명하고 쉬운 용어로 대체하세요.

3. 모르면 모른다고 하세요. 모든 답변은 명확한 근거를 가진 사실에 기반합니다. 예) 질문을 이해하기 어렵습니다. 잘 모르겠습니다.

4. 참조문서와 참조뉴스 기반으로 사용자의 질문에 틀린 내용이 있다면 수정해서 답변하세요.

5. 사용자의 개인정보 및 민감정보(주민등록번호, 카드정보 등)가 질문에 들어오면 '개인의 민감한 정보를 입력하지 마세요!'라는 경고를 띄우고 답변 생성하는 것을 멈춥니다.

6. 답변 형식에 맞추세요.

이전 메시지: {chat_history}

질문: {question}

참조문서: {context_pdf}

답변 형식:

🍀 머니또님의 투자성향은 [...]입니다.

추천하는 투자상품은 ... 입니다.

1. 투자상품:

- 특징 및 리스크:

2. 투자상품:

- 특징 및 리스크:

...

"""

- c. 프롬프트 예시 - 사용자의 투자 성향과 원하는 투자 상품, 질문에 관한 프롬프트

```
recommend_template = f"""
```

당신은 금융 상담 전문가이며, 고객의 질문에 맞는 대답을 하고 이에 대한 근거를 참조문서에 기반해 대답하는 역할을 맡고 있습니다.

고객이 원하는 금융투자상품은 {choice}입니다.

""" + """

1. 참조문서를 참고해서 사용자가 원하는 금융투자상품과 경제 지수 및 요인과의 관계를 파악하세요.

- 금융투자상품이 큰 수익성을 가지는 시점과 리스크를 가지는 시점을 분석하세요.

2. 1번 결과와 참조뉴스를 참고하여 사용자의 질문에 부합하는 답변을 하세요.

- 상품/종목 선택 기준은 명확하고 논리적이어야 하며, 장단점 및 리스크가 균형 있게 포함되도록 작성하세요.

- 추천 이유는 뉴스를 통해 현재 경기 상황을 파악하고 요약한 결과를 보고 최신 동향을 반영합니다. 이후 이득을 취할 것으로 예상되는 종목을 추천하고, 그 근거를 참조뉴스에서 찾아 구체적으로 덧붙이세요.

- 만약, 사용자가 선택한 금융투자상품이 리스크를 가지는 시점이라면 안정적인 다른 투자상품(예적금, 채권, 펀드 등등)을 추천합니다.

3. 모르면 모른다고 하세요. 모든 답변은 명확한 근거를 가진 사실에 기반합니다. 예) 질문을 이해하기 어렵습니다. 잘 모르겠습니다.

4. 참조문서와 참조뉴스 기반으로 사용자의 질문에 틀린 내용이 있다면 수정해서 답변하세요.

5. 사용자의 개인정보 및 민감정보(주민등록번호, 카드정보 등)가 질문에 들어오면 '개인의 민감한 정보를 입력하지 마세요!!'라는 경고를 띄우고 답변 생성하는 것을 멈춥니다.

6. 답변 형식에 맞추세요.

이전 메시지: {chat_history}

질문: {question}

참조문서: {context_pdf}

참조뉴스: {context_news}

답변 형식:

- 현재 경제 상황:

- 선택하신 금융투자상품:

- 추천 종목:

- 종목 추천 이유:

"""

d. 답변 생성: GPT-4o-mini

RAG

파이프라인 평가 및 결과

평가 방법

- 정량 평가: **RAGAS** 평가 지표
 - **context_precision**: 검색한 문서 중에서 진짜로 관련된 문서가 차지하는 비율
 - **context_recall**: 실제로 관련된 문서 중에서 얼마나 많이 검색에 성공했는지
- 정성 평가
 - 샘플링 방식
 - 무작위로 10개의 질문을 선택하여 챗봇의 답변을 평가합니다.
 - 질문은 제품 매뉴얼의 다양한 섹션에서 추출된 내용을 기반으로 구성합니다.
 - 평가 항목
 - 정확성: 생성된 답변이 제품 매뉴얼의 내용과 얼마나 일치하는가?
 - 관련성: 답변이 검색된 문서/데이터와 관련이 있는가?
 - 명확성: 답변이 쉽게 이해되고 논리적으로 명확한가?
 - 평가 절차
 - 각 질문에 대해 생성된 답변을 매뉴얼의 실제 내용을 기준으로 비교 검토합니다.
 - 관련성이 낮거나 잘못된 답변은 피드백을 기록하여 개선 방안을 도출합니다.
 - 평가 결과를 바탕으로 생성된 답변의 장단점을 정리하고, 추가 최적화 방향을 제시합니다.
 - 평가 질문
 1. 현재 금리가 떨어지려고 하고 있어서 채권을 사고자 합니다. 어떤 부문의 회사채를 구입하는 것이 좋겠습니까?
 2. 나는 안정추구형을 투자를 하고 싶은데, 현재 경기 상황을 고려해봤을 때, 어떤 종목에 투자를 하는 게 좋을 것 같아? 그리고 장투를 하는게 좋을까 아니면 단투를 하는게 좋을까?
 3. 너가 CMA를 추천해줬는데 MMW도 괜찮다고 들었거든. 두개의 차이가 뭐야?
 4. 나는 인생은 한 방이라고 생각하는 공격투자형이 나왔어. 지금 트럼프가 내놓은 공약을 봤을 때 대통령에 취임하기 전까지 어떤 주식 종목이 단기간에 빠르게 오를지 분석해줘. 매수 시점과 매도 시점에 대해서도 알려주면 좋겠어.
 5. 내가 지금 마이크로소프트 주식을 투자했는데, 왜 언제 빠지는 게 좋을까?
 6. 나는 위험중립형 투자 성향을 가지고 있는데, 펀드에 집중해서 투자를 하고 싶어. 하지만 펀드를 잘 모르고 있기 때문에 현재 정세를 고려해서 어떤 펀드가 좋은지 알려줘.
 7. 바이오 산업군에 투자를 하고자 합니다. 주식이나 채권을 구입하였을 때, 수익을 얻을 수 있을 것으로 보이나요?
 8. 나는 안정추구형이 나와서 예금을 추천받았는데 좀 더 수익률이 높은 상품이 있었으면 좋겠어. 리스크를 약간 더 감수해도 되니까 다시 상품추천해줘.
 9. 주식에 투자할까 하는데 최근 미국과 중국 간의 관계는 어떤지 분석해서 최대 수혜를 받을 종목을 알려줄 수 있어?
 10. ETF에 투자하고 싶은데, 리스크가 높은 종목은 조금 무섭네. 최근 경기나 트렌드에 미루어봤을 때 안정적인 ETF는 어떤 건지, 근거와 예시를 들어서 설명해줘.

평가 결과

- 정량 평가

VectorDB	Retriever	Rerank	Precision	Recall	Error
Pinecone	Sparse	X	0	0.7331	
Pinecone	Dense	X	0.5	0.7220	
Pinecone	Dense	O	0.5	0.8854	
Pinecone	Ensemble	X	0	0.7120	

Pinecone	Ensemble	O	0	0.7750	maxToken Issue
----------	----------	---	---	--------	----------------

=> 마지막 적용 행에서 Ensemble + Rerank 을 도입했을 때의 성능이 3번째 적용 행인 Dense + Rerank 을 도입했을 때보다 더 높은 결과를 예상합니다. 그 이유는 금융이라는 특정 도메인에서 키워드 중심의 Sparse Retriever가 중요한 역할을 할 것이기 때문입니다. 하지만 같은 context_recall 값이 나온 이유는 Ensemble + Rerank 을 적용했을 때, maxToken 이라는 Upstage Solar Mini 모델에서 오류가 나와서 성능이 예상보다 낮게 나온 것 같습니다. maxToken의 경우도, chunk_size 문제로 파악했지만 적당한 값이 선정되었고, 해당 모델에 문제가 있다는 것을 강사님을 통해 전해듣게 되었습니다. 이에 경우, 추후 개선시킬 수 있는 부분입니다.

- 정성 평가
 - 9. 주식에 투자할까 하는데 최근 미국과 중국 간의 관계는 어떤지 분석해서 최대 수혜를 받을 종목을 알려줄 수 있어?

아 그렇구나. 주식에도 투자할까 하는데 최근 미국과 중국간의 관계는 어떤지 분석해서 최대 수혜를 받을 종목을 알려줄 수 있어?

- **현재 경제 상황**: 최근 미국과 중국 간의 관계는 긴장 상태에 있으며, 이는 글로벌 경제에 큰 영향을 미치고 있습니다. 미국은 중국의 자가 제품에 대한 관세를 부과하고 있으며, 중국의 경기 둔화와 부동산 시장의 부진이 한국을 포함한 여러 국가의 수출에 부정적인 영향을 미치고 있습니다. 그러나 미국은 AI 및 기술 산업에 대한 대규모 투자를 유지하고 있으며, 이는 미국 주식 시장의 긍정적인 전망을 뒷받침하고 있습니다.

- **선택하신 금융투자상품**: 주식

- **추천 종목**: NVIDIA, Tesla, Microsoft

- **종목 추천 이유**: AI와 관련된 기술주들은 현재 미국 시장에서 가장 큰 수혜를 받을 것으로 예상됩니다. NVIDIA는 AI 칩 제조업체로, AI 기술의 발전과 함께 수요가 급증하고 있습니다. Tesla는 전기차 및 자율주행 기술에서 선두주자로, 미국의 친환경 정책과 맞물려 성장 가능성이 큼니다. Microsoft는 클라우드 서비스와 AI 솔루션을 통해 지속적인 성장을 기대할 수 있습니다. 이러한 기업들은 미국의 기술 중심 성장과 AI 혁명에 힘입어 주가 상승이 예상되며, 중장기적으로 안정적인 수익을 기대할 수 있습니다.

다만, 미국과 중국 간의 무역 갈등, 인플레이션 재반등, 정치적 리스크 등은 주식 투자 시 고려해야 할 리스크 요인입니다.

- 해당 종목 및 정세와 관련된 최신 뉴스 기사 검색 후 답변을 제공합니다.

⇒ 챗봇이 무작위로 선정된 10개의 정성평가 질문에 대해 정확성, 관련성, 명확성을 갖춘 답변을 잘 생성했습니다. 예시로 사용한 위 질문의 경우 미국과 중국 간 정세에 대한 최신 뉴스를 검색한 뒤, 긴장 상태에 대한 원인 및 향후 전망, 추천 종목 및 이유를 제시하고 있습니다. 해당 답변은 템플릿을 잘 유지하고 있으며, 검색된 뉴스와 충분히 관련있고, 논리적입니다. 따라서 챗봇의 모든 답변이 정성평가 기준을 통과했습니다.

<p>결론 및</p> <p>향후 발전 방향</p>	<p>결론</p> <p>- 설계된 RAG 파이프라인은 벡터 데이터베이스(Pinecone)을 활용했습니다. Retrieve와 Embedding 과정에서 관련된 정보를 빠르게 추출하고, CrossEncoderReranker로 이를 최적화하여 정확한 답변을 제공했습니다.</p> <p>- LangGraph를 활용하여 사용자 thread_id와 메시지 내용을 연결하여 개인의 “투자 성향”과 “투자 니즈”에 맞춘 투자 상담을 할 수 있도록 하였습니다.</p> <p>- 이 과정을 통해 사용자에게 빠르고 직관적인 투자 상품 및 종목을 추천했으며, 최신 뉴스 기사를 통해 다양한 금융 이슈에 대해서 적절하게 대응해낼 수 있었습니다.</p> <p>향후 발전 방향</p> <ul style="list-style-type: none">- 실시간 증시, 물가지수 등 경제 상황 분석과 투자 종목 선택에 영향을 미치는 데이터 추가- 최신 뉴스를 크롤링하여 Vector DB에 저장하는 시간을 줄여 사용자에게 더 빠르게 정보 제공- 회원 관리 DB를 만들어 개인 투자 성향과 종목을 저장하여 최적의 투자 전략을 제공- 투자 시장의 변화에 따른 모델 성능 변화를 지속적으로 모니터링하여 데이터 업데이트가 필요할 경우 신속하게 반영할 수 있도록 유지보수
-----------------------------	--