4차 회의 (3/31)

▼ (강병진) 모든 회사 리스트 제공 + 섹터 나눠서 회사 리스트업 + 검색 기능

1. 기능, 특징

- 기능
 - 。 전체 기업 리스트 조회
 - 。 섹터별 필터 기능
 - 。 키워드 기반 기업 검색
- 특징
 - RESTful API 기반의 백엔드 서버(Spring Boot)
 - 。 JSON 포맷으로 클라이언트에 데이터 제공
 - 。 빠른 검색을 위한 DB 인덱싱 적용

2. 최종 설계 결과물

- 형태: 백엔드 API 서버 (Spring Boot 기반 소프트웨어 프로토타입)
- 외적 형태: 프론트엔드에서 호출 가능한 기업 정보 API 제공

3. 시스템 구성과 기능

- ◆ 대표 기능: 회사 리스트 제공
- 하위 기능:
 - o /api/companies : 전체 조회
 - /api/companies/search?keyword=삼성 : 키워드 포함 기업 검색
 - o /api/companies/sector?sector=IT : 섹터 필터
- 달성 수준:
 - 。 필터링 및 검색 정확도 100%
 - 。 빠른 API 응답 속도

4. 다이어그램

☑ 유스케이스 다이어그램

- 사용자는 "기업 전체 보기", "섹터별 보기", "검색" 3가지 기능 사용
- 시스템은 해당 요청에 따라 필터링된 데이터를 응답

☑ 시스템 블록다이어그램

[사용자 요청]

 \downarrow

[프론트엔드 화면 (검색 입력, 버튼 클릭)]

↓ HTTP 요청

[Spring Boot API 서버]

↓ JPA Repository 호출

[MySQL DB]

 \downarrow

[검색 결과 반환 → 프론트]

☑ 시퀀스 다이어그램 (검색 예시)

사용자 → 프론트엔드: "삼성" 검색

프론트 → API 서버: GET /api/companies/search?keyword=삼성

API 서버 → DB: SELECT * FROM company WHERE name LIKE '%삼성%'

DB → API 서버: 검색 결과 반환

API 서버 → 프론트: JSON 응답 반환 프론트 → 사용자: 검색 결과 표시

√ (3) 대안 도출 및 구현 계획

1. 대안 비교

대안	장점	단점
모든 데이터 프론트에서 필터	빠름, 서버 부담 적음	데이터 양 많으면 느림
서버에서 필터 및 검색 (선택안)	대용량 처리 유리	서버 성능 요구
Open API 활용	실시간 데이터 확보 가능	의존성 큼, 과금 가능성

→ **서버 측 필터/검색 처리 방식 선택** (검색 속도 및 대용량 확장성 확보)

2. 구현 계획

단계	내용
1	MySQL에 기업 리스트 데이터 입력
2	Spring Boot로 API 서버 구성 (JPA 기반)
3	전체 조회, 섹터 필터, 검색 기능 구현
4	프론트와 연결 및 테스트
5	성능 최적화 (인덱스, 캐시 등)

3. 자료 구조 / 알고리즘

- 자료구조: RDB 기반 Table (회사명, 섹터, 종목코드 등)
- 알고리즘:
 - 검색 시 LIKE 쿼리 + 인덱싱
 - 섹터는 ENUM 또는 TEXT 필터링

4. 플로우차트 (검색 흐름)

```
[사용자 검색 입력]

↓

[프론트엔드 GET 요청]

↓

[Spring Boot 컨트롤러]

↓

[서비스 → JPA 쿼리 실행]

↓

[DB 검색 결과 반환]

↓

[서비스 → 컨트롤러 → 프론트로 응답]
```

▼ (서희정) 기본적인 교육 (재무제표 용어) 튜토리얼 + 링크 연결

1. 소개 (Introduction)

- 삼성전자라는 실제 기업을 예시로 사용.
- 재무제표 분석의 목표: 삼성전자의 재무 상태, 수익성, 현금 흐름을 이해하기.

• 사용 도구: Python, Open DART API, React (웹 인터페이스)

▦ 모바일 재무제표 분석 웹 앱 구조

🔍 기능 개요

- 1. 특정 기업 선택 페이지
- 2. 탭 메뉴 (재무제표, 손익계산서, 현금흐름표, Valuation, 총평)
- 3. 각 탭마다 해당 기업의 지표를 분석하여 텍스트와 이미지로 표현
- 4. 하단에는 데이터베이스(DB), API 연결, 레이아웃 조정 기능 등 포함
- React로 **탭 컴포넌트**를 구현하여 각 메뉴를 전환하며 확인 가능
- Flask로 DART API에서 데이터 불러오기
- 각 분석 결과를 텍스트와 그래프로 표현

🕎 React + Flask (Python)로 구현하기

- Frontend (React): 사용자가 볼 수 있는 화면을 구현
- Backend (Flask): Open DART API로 데이터를 불러오고 분석하는 기능 제공
- Database (MySQL): 사용자 데이터와 분석 결과를 저장

📕 React 구성 요소

- 1. 메인 페이지 (기업 선택 페이지)
- 2. 탭 메뉴 (Tabs Component)
- 3. 분석 결과 출력 컴포넌트 (Text Analysis, Graph Visualization)

구현 예시

```
// App.jsx
import React, { useState } from 'react';
import { Tabs, Tab } from "@/components/ui/tabs";
import AnalysisTab from './components/AnalysisTab';
const App = () ⇒ {
```

AnalysisTab.jsx (각 탭별 내용 렌더링)

```
// AnalysisTab.jsx import React from 'react';

const AnalysisTab = ({ activeTab }) → {
  const content = {
  "재무제표": "삼성전자의 최근 재무제표 데이터입니다. 자산, 부채, 자본을 분석합니" 손익계산서": "삼성전자의 매출, 영업이익, 순이익을 분석합니다.",
  "현금흐름표": "삼성전자의 영업활동, 투자활동, 재무활동의 현금흐름을 분석합니다" Valuation": "삼성전자의 가치 평가. PER, PBR 등 주요 지표 분석.",
  "총평": "삼성전자의 종합 평가: 최근 매출이 상대적으로 줄었어요! (최근 데이터 침 };

return (
  <div className="mt-4 p-4 border rounded shadow-sm">
  <h2 className="text-xl mb-2">{activeTab}</h2>
```

```
{content[activeTab]}
</div>
);
};
export default AnalysisTab;
```

Flask 백엔드 (Python) 예시

```
from flask import Flask, request, jsonify import requests

app = Flask(__name__)

@app.route('/get_data', methods=['GET'])
def get_data():
    corp_code = request.args.get('corp_code')
    api_key = "YOUR_API_KEY"
    url = f"https://opendart.fss.or.kr/api/fnlttSinglAcnt.json?crtfc_key={api_kresponse = requests.get(url)}
    data = response.json()
    return jsonify(data)

if __name__ == '__main__':
    app.run(debug=True)
```

- ▼ (서희정) 논의하고 싶은 부분
 - 1. pc로 하면 fnguide랑 다를 바가 없는 것 같음. 모바일 웹으로 만들어보는 건 어떤 가여? 요즘 주식도 핸드폰 mts로 하니까요!
 - 2. 청소년을 위한 재무제표 교육 프로그램이었는데, 어째 조사를 하면 할수록 주식 투자를 위한 재무분석 DB를 만들게 되는 느낌입니다. 그래서 정체성 혼란이 오고 있어요ㅜ 청소년이라 해서 기본적인 것만 가르키자니 뭔가 이상해지는 것 같아요 차라리 그냥 교육프로그램으로 하고 난이도를 기초 용어에서 기본필수적인 건 다 알려주는 부분으로 하는 건 어떠신가용
 - 3. 프로그램 이름 : 튜자 (투자에 1을 더해 튜자ㅋㅋㅋㅋㅋ)
- ▼ (서희정) 재무제표 분석 지표 추가 (by 경영학과 교수님 조언 : 아무리 그래도 PER, PBR은 봐야지)

1. 기업의 건강 상태 파악 (재무 안정성 분석) 지표

- 유동성 비율: 단기 부채를 갚을 수 있는 능력을 측정.
 - **유동비율 (Current Ratio)** = 유동자산 / 유동부채 (100% 이상이 바람직)
 - 당좌비율 (Quick Ratio) = (유동자산 재고자산) / 유동부채 (당좌비율이 높을 수록 좋음)
- **부채비율 (Debt Ratio)**: 회사의 재정 구조 안정성을 평가.
 - 부채비율 = 총부채 / 총자본 (100% 이하가 안정적)
 - 。 업종마다 다르지만, 너무 높은 부채비율은 위험 신호
- 이자보상비율 (Interest Coverage Ratio): 영업이익으로 이자 비용을 얼마나 감
 당할 수 있는지 보여줌.
 - **이자보상비율** = 영업이익 / 이자비용 (200% 이상이면 안전)

2. 성장 가능성 평가 (성장&수익성 분석) 지표

- 매출 성장률 (Revenue Growth Rate): 매출이 얼마나 증가하고 있는지 평가.
 - 매출 성장률 = [(금년도 매출 전년도 매출) / 전년도 매출] × 100
 - 꾸준히 증가하는 것이 좋지만, 지나치게 빠르면 지속 가능성을 의심.
 - ▼ 판단 기준
 - (급격한 성장의 배경이 **비정상적인 원인**일 수 있음. ex) 일회성 이벤트, 정부 보조금, 혹은 지나치게 공격적인 마케팅 등)
 - / 매출 성장률 판단의 **일반적인 기준**:
 - 연간 성장률이 **20~30% 이상**이면 빠른 성장으로 간주될 수 있어요.
 - 산업 평균 성장률보다 **2배 이상 높을 때**도 의심해볼 필요가 있음.
 - Ⅰ 경우에 따른 해석:
 - 스타트업이나 신산업: 초반에 빠르게 성장하는 게 자연스럽기 때문에
 50% 이상의 성장률도 괜찮을 수 있어요.
 - **성숙한 산업**(예: 전통 제조업): 연 10% 이상의 성장은 비정상적으로 빠를 수 있음.
- 영업이익률 (Operating Profit Margin): 매출에서 영업비용을 제외한 영업이익의 비율.

- 。 영업이익률 = 영업이익 / 매출 x 100
- 산업 평균보다 높은 것이 좋음.
- ▼ 산업 평균 계산 방법:
 - 1. 동일 산업 내 주요 기업들의 영업이익률 수집

예를 들어, 반도체 산업이라면 삼성전자, SK하이닉스 등 주요 기업의 영업 이익률을 모두 조사.

2. **평균 값 산출 (산술 평균)**산업 평균 영업이익률=기업 수개별 기업들의 영업 이익률 총합

산업 평균 영업이익률 = 개별 기업들의 영업이익률 총합기업 수

- 3. **중간값 사용 (Median)** 극단적인 값(예: 아주 높은 이익률)을 제외하고 판단하고 싶을 때 사용.
- 🔍 데이터 출처:
 - Open DART API (국내 기업), Yahoo Finance, Reuters 등에서 산 업별 주요 기업의 데이터를 수집하여 계산.
- ROE (Return on Equity, 자기자본이익률): 주주의 투자 대비 수익률을 측정.
 - ROE = 순이익 / 자기자본 × 100
 - 。 일반적으로 10% 이상이면 우수한 것으로 평가돼.

3. 투자 리스크 평가 (안정성 분석)

- 현금흐름 분석 (Cash Flow Analysis): 기업의 실제 현금 창출 능력을 평가.
 - 。 **영업활동 현금흐름**이 지속적으로 양수인지 확인.
 - ▼ 영업 활동 현금 흐름
 - : 회사가 본업을 통해 벌어들인 실제 현금 흐름

영업활동 현금흐름 = 순이익 + 비현금 비용(감가상각비) - 운전자본의 변화

- **순이익 (Net Income)**: 손익계산서에서 가져옴.
- 비현금 비용: 감가상각비, 무형자산 상각비 등 (현금 지출이 없는 비용).

• 운전자본의 변화:

- 。 운전자본 = 유동자산 유동부채
- 증가하면 현금 유출(자산 증가), 감소하면 현금 유입(자산 감소).
- **잉여현금흐름 (Free Cash Flow)** = 영업활동 현금흐름 자본지출(CAPEX)
- 。 잉여현금흐름이 플러스면 긍정적.
- 부채 상환 능력: 회사의 현금흐름으로 부채를 갚을 수 있는지 평가.
 - **부채비율**, **이자보상비율**, 그리고 **자기자본비율**을 함께 분석.
 - 자기 자본 비율 : 높을수록 안정적이며, 50% 이상이면 좋음.

자기자본비율 =
$$\frac{$$
자기자본 $}{$ 총자산 $}$ \times 100

▼ 분석 방법

- 부채비율이 높더라도 이자보상비율이 충분히 높으면 안전할 수 있음.
- 자기자본비율이 높으면 부채비율이 높아도 위험성이 낮음.
- 세 가지를 함께 분석하여 안정성과 위험성을 동시에 고려해야 함.
- 산업 및 시장 리스크: 외부 요인도 고려해야 함. (정책 변화, 경쟁 상황, 기술 변화 등)
 - : 웹스크래핑으로 산업 및 시장 리스트 뉴스들 가져오

4. 적정 주가 평가 (Valuation, 가치 평가)

- PER (Price to Earnings Ratio, 주가수익비율):
 - PER = 주가 / 주당순이익(EPS)
 - 낮을수록 저평가된 것으로 볼 수 있지만, 성장주의 경우 높은 PER도 ㄱㅊ
- PBR (Price to Book Ratio, 주가순자산비율):
 - **PBR** = 주가 / 주당순자산(BPS)
 - 。 PBR ≤1: 자산 가치보다 저평가된 것일 가능성이 있음.
- EV/EBITDA (Enterprise Value to EBITDA Ratio) :

• 기업 가치(EV)를 EBITDA로 나눈 값으로, 이 수치가 낮을수록 상대적으로 저평가 된 것일 수 있음.

▼ 계산 방법

EV (Enterprise Value, 기업 가치): 회사가 인수될 때 실제로 지불해야 하는 총 금액.

$$EV = 1$$
시가총액 + 총부채 - 현금 및 현금성 자산

EBITDA: 기업의 본업에서 발생하는 수익을 측정하는 지표 (비용 구조를 비교할 때 유용).

EBITDA = 영업이익 + 감가상각비 + 무형자산상각비

DCF (Discounted Cash Flow, 할인 현금흐름 모델):

- 미래의 현금흐름을 현재 가치로 환산하여 기업의 가치를 평가.
- 가장 정교하지만, 많은 가정이 필요함.
- ▼ DCF 계산 방법
 - DCF는 미래의 현금흐름을 현재 가치로 환산하여 기업의 가치를 평가하는 방법

$$ext{DCF} = \sum_{t=1}^n rac{FCF_t}{(1+r)^t} + rac{TV}{(1+r)^n}$$

- FCFt : 특정 연도의 **잉여현금흐름 (Free Cash Flow)**.
- r: 할인율 (일반적으로 WACC, 가중평균자본비용).
- n: 평가 기간 (예: 5년 또는 10년).
- TV: Terminal Value, 평가 기간 이후의 가치를 추정.

- DCF 분석에서 중요한 것:
 - 정확한 FCF 예측과 적절한 **할인율 설정**이 필요.
 - 。 터미널 밸류 계산 방식도 여러 가지가 있음 (Gordon Growth Model 등).

5. 의사결정의 근거 제공 (종합 분석)

모든 지표를 종합하여 투자 결정을 내릴 때 고려할 점:

- 종합 점수화 또는 랭킹화: AHP, TOPSIS 같은 다기준 의사결정 기법을 사용해서 여러 지표를 통합할 수도 있음.
- 업종 평균과의 비교: 같은 산업 내 다른 기업들과의 비교가 중요함.
- 트렌드 분석: 단일 시점의 데이터보다는 여러 기간의 추세를 보는 것이 더 의미 있음.
- 정성적 요소의 고려: 브랜드 가치, 기술 혁신, 경영진의 능력 등도 분석에 포함할 필요가 있음.
- ▼ (서희정) 재무상태표, 손익 계산서, 현금 흐름표에서 우리가 추출할 항목들

1. 재무상태표 (Balance Sheet)

🏲 추출할 항목들

- 유동자산 (Current Assets)
 - 예: 현금 및 현금성 자산, 매출채권, 재고자산, 기타유동자산 (미수금, 선급비용, 유동금융자산)
- 유동부채 (Current Liabilities)
 - 。 예: 매입채무, 단기차입금, 미지급비용, 선수금, 유동금융부채
- 총부채 (Total Liabilities)
 - 예: 유동부채 + 비유동부채 (장기차입금, 사채, 미지급법인세, 확정급여부채)
- 자기자본 (Total Equity)
 - 예: 자본금, 자본잉여금, 이익잉여금, 기타포괄손익누계액, 자사주
- 재고자산 (Inventory)
 - 유동자산의 일부로, 당좌비율 계산에 필요
- 현금 및 현금성 자산 (Cash and Cash Equivalents)

2. 손익계산서 (Income Statement)

🏲 추출할 항목들

- 매출액 (Revenue, Sales)
 - 예: 제품 판매, 서비스 제공, 로열티 수익, 투자수익 (지분법 이익, 배당금)
- 영업이익 (Operating Income)
 - 매출액 영업비용 (매출원가, 판매비, 관리비 등)
- 순이익 (Net Income)
 - 최종적인 이익으로, 모든 비용(영업비용, 이자비용, 세금 등)을 제외한 금액
- 이자비용 (Interest Expense)
 - 금융 비용으로, 손익계산서 주석에 포함될 수 있음
- 감가상각비 (Depreciation)
 - 。 비유동자산의 감가상각에 따른 비용
- 무형자산 상각비 (Amortization)
 - 。 무형자산의 가치 감소에 따른 비용

3. 현금흐름표 (Cash Flow Statement)

🏲 추출할 항목들

- 영업활동 현금흐름 (Operating Cash Flow)
 - 본업을 통해 발생한 현금흐름 (예: 매출 현금 유입 영업비용 현금 유출)
- 잉여현금흐름 (Free Cash Flow, FCF)
 - 。 영업활동 현금흐름 자본지출 (CAPEX)
- 자본지출 (CAPEX)
 - 。 설비 투자, 건물 등 비유동자산의 취득 비용
- 재무활동 현금흐름 (Financing Cash Flow)
 - 。 예: 배당금 지급, 차입금 변동, 주식 발행, 자사주 매입 및 처분
- 투자활동 현금흐름 (Investing Cash Flow)
 - 。 예: 설비 투자, 자산 매각, 유가증권 매매, 비유동자산 취득

4. 외부 데이터

- 발행주식수 (Shares Outstanding)
- 주가 (Stock Price)
- 시가총액 (Market Capitalization) = 주가 × 발행주식수
- 할인율 (Discount Rate, r)

▼ (서희정) 선행기술 관련 논문

초등사회과 경제교육에서 금융이해력 향상을 위한 SEC 금융교육 프로그램 개발 및 적용 https://www.dbpia.co.kr/journal/detail?nodeld=T16854898

다문화가정 아동을 위한 온라인 금융교육프로그램 개발

https://www.dbpia.co.kr/journal/articleDetail?nodeld=NODE11633844

가치기반 아동금융교육 프로그램 개발 및 효과평가 : 개인재무관리 영역별 가치를 중심으로

https://www.dbpia.co.kr/journal/detail?nodeld=T15827669

검색

https://www.dbpia.co.kr/search/topSearch?searchOption=all&query=금융교육+프로그램+개발

노인소비자 대상 소비자금융 교육프로그램 개발

https://www.dbpia.co.kr/journal/detail?nodeld=T13991454

▼ (이설후) 챗봇구현

2. 구현 방법

2.1. 텍스트 추출 & 전처리

- 1. 드래그 & 드롭 혹은 텍스트 입력 이벤트
 - 브라우저 상에서 사용자가 텍스트를 드래그한 뒤, 선택한 영역을 "드롭존" 혹은 "채팅 입력창" 등에 놓으면 해당 텍스트를 서버 또는 클라이언트 측 스크립트로 전달받는다.
 - 예: React에서 onDrop 이벤트, JavaScript FileReader 등을 활용 가능.

2. 전처리 및 필터링

- 사용자가 드래그해 넣은 텍스트가 재무제표 정보인지, 회사 소개인지 등 간단한 분류를 할 수 있음.
- 텍스트 길이가 너무 길면 요약(summarization) 과정을 거쳐 "챗봇이 이해하기 적절한 길이"로 축소.
- 재무제표 숫자나 재무 지표, 전문 용어 등은 **토큰화** 또는 **키워드 추출** 기술로 걸러내 별도의 필드로 보관하면 이후 AI 설명 시 더욱 정확한 설명이 가능해짐.

2.2. AI 모델 연동

1. 사전 학습된 대규모 언어 모델(LLM) 활용

- 예: OpenAI의 GPT-3.5 / GPT-4, 혹은 Hugging Face Transformers 기반 모델 등.
- 직접 모델을 호스팅할 수도 있지만, 초기에 빠른 프로토타입을 위해서는 API 기반 접근(예: OpenAl API)이 훨씬 편리함.

2. 프롬프트 엔지니어링(Prompt Engineering)

- 사용자 입력 텍스트와 함께, "이 텍스트를 초심자/청소년 대상 금융 교육용으로 쉽게 풀어서 설명해 달라", "추가적으로 적절한 예시를 들어 달라" 등의 **지시 (prompt)**를 함께 전달한다.
- 지표 해석이라면, "이 텍스트 내에서 ROE, PER, PBR 등 용어가 등장하면 간 단하고 쉬운 예시를 들어달라"와 같은 시스템 메시지를 미리 구성할 수 있음.
- 꼬리질문을 유도하기 위해, 예: "궁금한 부분이 있으면 질문을 유도해 달라"는 식의 프롬프트를 추가한다.

3. 체크포인트 / 파인튜닝(Fine-tuning) / 임베딩(Embedding) & 벡터 DB

- (선택) 프로젝트에서 사용하는 용어 정의나 회사별 재무정보 등 도메인 데이터 가 풍부하다면, 이를 **벡터 DB(예: Pinecone, Weaviate, Milvus 등)**에 임베딩 형태로 저장해 두고, 사용자가 드래그한 내용과 유사도가 높은 자료를 검색 후 모델에 전달("Retrieval-Augmented Generation(RAG)" 방식).
- 이렇게 하면 **도메인 특화된** 정확한 답변을 생성할 수 있음.
- 또는 Hugging Face 모델을 **파인튜닝**하여 반복 질문에도 안정적으로 답변하 도록 학습 가능(단, 데이터 준비와 리소스가 필요).

2.3. 대화 플로우 설계

1. 사용자 텍스트 입력 → AI 응답

AI 응답을 받은 후, 클라이언트(웹/앱)에서 해당 응답을 표시.

• 이때 단순히 "답변"만 주는 것이 아니라 '계속 질문하기' 혹은 '좀 더 쉽게 설명 해 달라' 같은 버튼을 제공해, 유저가 2차, 3차 꼬리질문을 던질 수 있도록 함.

2. 대화 컨텍스트 유지

- 꼬리질문에 대해 AI가 맥락을 유지하도록 하기 위해, 프롬프트에 "이전 대화 내용" 또는 "세션 히스토리"를 포함.
- 대화형 API 사용 시, **대화 히스토리**를 일정 길이(토큰 제한)에 맞춰 관리해야 함(너무 길면 API 호출 비용 및 처리 속도 문제).

3. 사용자 난이도/레벨 조절

- 예: 초등학생, 중학생, 고등학생 레벨별로 다른 난이도의 단어를 사용하도록 모델에 지시 가능.
- "설명을 쉽게 / 더 깊게 / 수학 공식까지 상세히" 등 선택지를 제공하면, 모델에 적절한 파라미터(temperature, max_tokens 등)나 프롬프트를 달리 줄 수도 있음.

3. 사용 기술 및 스택

3.1. 프론트엔드

- React / Vue / Angular 중 하나
- 드래그 & 드롭 기능 구현 시 React-Beautiful-DND 또는 HTML5 drag/drop API 사용 가능
- 채팅 인터페이스 구성 시 React-Chat-Widget, ChatGPT 스타일 UI 라이브러리 등 활용 가능

3.2. 백엔드

- Node.js(Express, NestJS 등) 또는 Python(FastAPI, Django) 등으로 API 서 버 구성
- AI 모델 연동(OpenAI API 등) 기능 구현
- 사용자 인증, 권한 관리(만약 청소년 및 교사 계정이 있다면), 데이터베이스 연동

3.3. AI 모델 연동

1. OpenAl API

• text-davinci-003, gpt-3.5-turbo, gpt-4 등의 엔진 사용

• API 키 관리, 프롬프트 엔지니어링, 토큰 관리(요금 계산)

2. Hugging Face + Transformers

- 직접 모델 호스팅 시, GPU/메모리 자원 필요
- Flask나 FastAPI로 Serving 구축 후 API로 제공 가능

3. LangChain / Haystack 등(선택)

- Retrieval-Augmented Generation 파이프라인 편리하게 구성
- 벡터 DB와 AI 모델의 질의응답 연결

3.4. 데이터베이스

- SQL: MySQL, PostgreSQL 등 -> 계정 관리, 히스토리 저장
- NoSQL: MongoDB 등 -> 대화 히스토리나 세션 데이터 기록
- **벡터 DB**(선택): Pinecone, Weaviate 등 -> 회사 재무제표 정보, 용어 사전 등 대규모 텍스트를 임베딩 형태로 저장 후 검색

4. 구현 난이도 & 고려사항

1. 난이도(중 ~ 상)

- 단순 OpenAl API 연동으로 "질문 → 답변" 정도의 MVP(최소 기능 제품)는 구현 난이도가 중간 정도.
- 그러나 **정확도**(특히 금융, 재무 관련)는 도메인 지식과 정확한 데이터 업데이트 가 필수.
- "꼬리질문" 시 맥락 유지와 대화 히스토리 관리가 필요해 **프롬프트 엔지니어링** 과 **대화 세션 설계**가 중요함.

2. 모델 선택

- SaaS(Software as a Service) 형태로 OpenAl를 쓰면 GPU 인프라 없이 손쉽게 시작 가능하지만, API 비용과 데이터 보안(개인정보, 민감 데이터)에 유의.
- 사내에 GPU 인프라가 있고, 전문 인력이 있다면 Hugging Face 모델 호스팅 고려 가능(초기 세팅 난이도 높음).

3. 데이터 보강

• 재무제표, 투자 용어, 회사별 정보 등 국내 상황에 맞는 **한글 데이터**를 충분히 준비해야 함.

• AI가 틀린 정보를 "헛소리(Hallucination)"로 말할 수 있으므로, RAG(데이터 검색 → 답변 생성) 방식을 사용하면 정확도를 높일 수 있음.

4. 사용자 경험(UX)

- 청소년이 사용하는 만큼 **UI는 직관적으로**, **용어는 쉽게** 설명해야 함.
- 특정 용어를 클릭하면 해당 용어의 정의로 이동할 수 있는 링크나 팝업 기능 등을 챗봇 대화창과 유기적으로 연결해주면 학습 효율이 올라감.

5. 추가 기능

- **토론 모드**: AI와 1:1로만 질문/답변이 아니라, 다른 학생/사용자도 함께 볼 수 있는 공개 토론 방식을 지원할 수도 있음(추가 난이도).
- 학습 이력 저장: 사용자가 질문한 내용, 봤던 답변들을 저장하고, 나중에 재접속 시 이어서 학습 가능하게 함.

5. 요약 및 제안 로드맵

1. 1차 프로토타입

- 프론트엔드(React 등)에서 텍스트 드래그 & 드롭 → 백엔드(Node/Python)
 에서 OpenAl API 호출 → 답변 표시
- 세션/대화 컨텍스트는 간단히 구현

2. 2차 고도화

- 벡터 DB 사용해 도메인 지식(재무제표 용어집, 회사 데이터) 통합
- 한글 투자 용어, 국내 재무제표 분석 특화

3. 3차 확장

- 꼬리질문, 난이도 조절, 사용자별 수준 맞춤형 답변
- UI/UX 최적화(링크로 상세 설명 이동, 퀴즈 기능, 학습 수준 평가 등)

▼ # 시각화 자동화 코드

import requests import pandas as pd import matplotlib.pyplot as plt

def fetch_dart_data(api_key, corp_code, year, report_code="11011"):
""" OpenDART API에서 데이터를 받아오는 함수 """
url = "https://opendart.fss.or.kr/api/fnlttCmpnyIndx.json"

```
params = {
    "crtfc_key": api_key,
    "corp_code": corp_code,
    "bsns_year": year,
    "reprt_code": report_code
  }
  response = requests.get(url, params=params)
  if response.status_code == 200:
    return response.json()
  else:
    print(f"Error {response.status_code}: {response.text}")
    return None
def json_to_dataframe(json_data):
  """ JSON 데이터를 Pandas DataFrame으로 변환하는 함수 """
  if json_data and "list" in json_data:
    df = pd.DataFrame(json_data["list"])
    return df
  else:
    print("No data available")
    return None
def plot_data(df, x_col, y_col):
  """ DataFrame 데이터를 바탕으로 그래프 생성하는 함수 """
  if df is not None and x_col in df.columns and y_col in df.columns:
    plt.figure(figsize=(10, 5))
    plt.plot(df[x_col], df[y_col], marker='o', linestyle='-', color='b',
label=y_col)
    plt.xlabel(x_col)
    plt.ylabel(y_col)
    plt.title(f"{y_col} Trend")
    plt.legend()
    plt.xticks(rotation=45)
    plt.grid()
    plt.show()
  else:
```

print("Invalid columns for plotting") # 🔷 실행 예제 API_KEY = "YOUR_API_KEY" CORP_CODE = "YOUR_CORP_CODE" YEAR = "2023" # 1 API 데이터 가져오기 json_data = fetch_dart_data(API_KEY, CORP_CODE, YEAR) # 🔼 DataFrame 변환 df = ison_to_dataframe(ison_data) # ③ 그래프 생성 (예시: 'thstrm_amount'을 'si_div' 기준으로) if df is not None: plot_data(df, x_col="sj_div", y_col="thstrm_amount") ▼ # 브리핑 코드 import requests import pandas as pd import matplotlib.pyplot as plt from transformers import pipeline def fetch_dart_data(api_key, corp_code, year, report_code="11011"): """ OpenDART API에서 데이터를 받아오는 함수 """ url = "https://opendart.fss.or.kr/api/fnlttCmpnyIndx.json" params = { "crtfc_key": api_key, "corp_code": corp_code, "bsns_year": year, "reprt_code": report_code } response = requests.get(url, params=params) if response.status_code == 200:

4차 회의 (3/31) 19

print(f"Error {response.status_code}: {response.text}")

return response.json()

else:

return None

```
def ison_to_dataframe(ison_data):
  """ JSON 데이터를 Pandas DataFrame으로 변환하는 함수 """
  if json_data and "list" in json_data:
    df = pd.DataFrame(json_data["list"])
    return df
  else:
    print("No data available")
    return None
def generate_summary(df):
  """ DataFrame 데이터를 기반으로 AI 브리핑 생성 """
  summarizer = pipeline("summarization")
  # 텍스트 요약할 재무 정보 정리
  financial_text = "\n".join(
    f"{row['si_div']}: {row['thstrm_amount']}원" for _, row in df.iterrows()
  )
  summary = summarizer(financial_text, max_length=100, min_length=30,
do_sample=False)
  return summary[0]['summary_text']
def generate_report(api_key, corp_code, year):
  """ 전체 프로세스 실행 및 보고서 생성 """
 json_data = fetch_dart_data(api_key, corp_code, year)
  df = json_to_dataframe(json_data)
  if df is not None:
    summary = generate_summary(df)
    print("\n 📌 **자동 브리핑 보고서** 📌")
    print(f" 가 사업 연도: {year}")
    print(df[['si_div', 'thstrm_amount']].to_string(index=False))
    print("\n 🥟 AI 요약 브리핑:")
    print(summary)
```

else:

print("**X** 데이터 없음")

실행 예제 API_KEY = "YOUR_API_KEY" CORP_CODE = "YOUR_CORP_CODE" YEAR = "2023"

generate_report(API_KEY, CORP_CODE, YEAR)

https://www.classting.com/

▼ (강병진) 철학, 정당성, 차별점

📌 프로젝트 철학

"청소년도 스스로 경제와 투자를 이해하고 판단할 수 있어야 한다."

기존의 금융교육은 주입식이거나 흥미를 유도하지 못하는 방식이 대부분이다. 우리는 청소년이 단순히 정보를 외우는 것이 아니라, 실생활과 연결된 금융 개념을 스스로 묻고, 이해하고, 판단해보는 경험을 하도록 돕고자 한다. 이를 위해 복잡한 UI나 용어 대신, 자연어로 대화하는 챗봇 인터페이스를 통해 쉽고 친숙한 방식으로 경제 개념과 재무지표를 설명받고 체험할 수 있도록 구성했다.

▼ 정당성

청소년 대상 금융교육은 많지만, 대부분 일방향적이고 파편화되어 있어 실질적인 이해로 이어지지 않는다.

강의 영상, 퀴즈, PDF 자료 등은 정보는 많지만 **맥락이 없고, 개인** 의 수준이나 궁금증에 따라 설명이 달라지지 않는다. 우리는 이를 보완하기 위해 AI 챗봇 구조를 적용해, 자연어 질문에 맞춤형으로 응답하고 실시간으로 지표를 분석해주는 상호작용 방식을 선택했다. 이

는 금융 개념을 스스로 탐색하고 연결하게 해주며, 사용자의 질문에 따라 유동적인 학습 흐름을 제공할 수 있다는 점에서 교육적 정당성이 충분하다.

🧠 차별점

단순한 챗봇, 단순한 금융교육 플랫폼이 아니다.

우리는 오픈소스 자연어 처리 모델(HuggingFace)을 활용해 사용 자의 문장에서 기업명, 지표명, 질문 의도를 파악하고, 내부 재무제표 데이터베이스와 연결해 실제 데이터를 기반으로 의미 있는 분석과 해석을 제공한다. 예를 들어 "삼성전자 PER이 18이면 괜찮은 거야?"라는 질문에, 챗봇은 업종 평균과 비교한 결과를 설명하며 단순한 정의가 아닌 판단 근거 중심의 응답을 제공한다. 이 구조는 기존금융교육의 한계를 극복하면서도, 기술적으로도 오픈소스를 적극활용해 확장성과 실용성을 갖춘다.

▼ ChatGPT랑 다른 점 (왜 이걸 따로 만드는가?)

1. 🞯 정확한 지표 해석 → 실제 데이터 기반

- ChatGPT는 개념은 잘 설명하지만,

 "삼성전자 PER이 18이면 어떤 상태냐?" 같은 질문엔 실제 지표를 조회해서 비교하
 진 않음
- → 우리 챗봇은 **기업 데이터 + 업종 평균**을 바탕으로 "고평가/저평가" 같은 **판단**을 해줌

2. 🧠 눈높이 + 설명 방식 통제 가능

- ChatGPT는 말투나 설명 수준이 고정되지 않음 (초등학생이 물어봐도 어려운 용어 쓸 수 있음)
- 우리는 말투, 설명 방식, 예시 등을 완전히 통제된 UX로 설계 가능

- GPT는 BlackBox임. 내부 로직/데이터 수정 불가
- 우리는 HuggingFace 모델, Spring 백엔드, 데이터 전부 우리가 컨트롤
 - → 사용자 피드백에 따라 모델 튜닝, 응답 방식 수정 등 **지속 개선 가능**

🦍 한 줄 요약

ChatGPT는 "똑똑한 백과사전"이고,

우리는 "진짜 재무데이터를 보고, 쉽게 설명해주는 튜터"임.

▼ (강병진) 기능

☑ "챗봇 + 재무제표 해석 시뮬레이션 (기초)"

【 단순 정보 전달 X →

"지표값 + 기준 비교" 기반으로

"좋다/나쁘다" 간단하게 말해주는 챗봇

◎ 예시 흐름

사용자: "삼성전자의 PER이 18인데 이거 괜찮아?"

출 챗봇: "동일 업종 평균 PER은 10이야. 삼성전자는 평균보다 높아서 고평가일 가능성이 있어."

🧩 기능 요약 (MVP 기준)

기능	설명
재무지표 DB	기업명, PER, PBR, ROE, 업종 평균 같은 정적 데이터

챗봇 UI	기본 입력창 + 응답창 (react-chat-ui 등)
사용자 질문 파싱	"PER", "ROE", "좋아?", "높아?" 등 키워드 추출
판단 로직	기준값과 비교해서 "좋다/나쁘다/보통"으로 응답 생성
챗봇 응답	"업종 평균보다 높음 → 고평가일 가능성 있음" 식 설명

💬 차별성 요약 (교수님 설득용)

- 단순 QnA 챗봇이 아니라 **지표 + 기준 비교**로 분석된 답변 제공
- 실제 데이터 기반의 맞춤형 피드백 구조
- UI/UX는 오픈소스 활용 → 챗봇 구조 설계에 집중
- 백엔드는 간단한 분석 로직이지만 정량 판단 알고리즘 존재
- ▼ (강병진) 토스증권이랑은 뭐가 다른가?

▼ [차이점]

1. 사용자 주도 vs 시스템 주도

- 토스: 지표를 보여주고, 시스템이 "이건 저평가예요"라고 자동 분석
- **너네**: 사용자가 "이게 괜찮은 거야?", "왜 고평가야?", "이럴 때 사도 돼?"라고 **직접** 질문
 - → 거기에 대해 자연어 해석 → 맞춤형 피드백 제공

2. 자연어 기반 인터랙션

- 토스는 정해진 틀의 정보 제공
- 너네는 사용자가 아무 질문이나 던져도 **의도 파악 + 지표 추출 + 의미 있는 판단 근** 거 제시
 - 。 예: "요즘 반도체 섹터 어때?" → 관련 업종 평균, 흐름, 대표 기업 PER 등 뽑아 서 답변
 - \circ 예: "LG에너지솔루션 비싸?" → 배터리 업종 비교, PER/PBR 해석, 성장성 코 멘트 등

3. 교육적 목적 강조 (판단력 훈련)

- 토스는 분석 보여주고 끝
- 너네는 **왜 그런 판단이 나왔는지 해석**, 더 나아가서

- 。 "이걸 보면 뭐가 중요한 포인트인지" 알려줌
- 투자 초보가 **스스로 투자 판단 능력을 키우는 구조**

4. 확장성 (사용자 맞춤형 시나리오 가능)

- 예:
 - 。 "나는 PER보다 성장률이 더 중요해" → 너네는 그 성향까지 반영 가능
 - 。 "이 회사 작년에 흑자전환했는데, 이런 경우 PER은 어떻게 해석해야 돼?" → 상황별 맥락 포함한 응답 가능

🔑 한줄 요약

토스는 정보를 보여주는 서비스,

너네는 정보를 "이해시키고 판단 도와주는" 서비스.

- ▼ (강병진) 다트포인트와 차별점
 - 대상 및 목적
 - o dartpoint.ai: 주로 재무제표와 공시 데이터를 기반으로 기업 분석 및 자동 보고서를 제공
 - 우리 프로젝트 (INVestigator): 단순 분석 보고서 제공을 넘어, 사용자가 직접 질문하고 해석하는 대화형 AI 멘토로서 투자 판단력 훈련에 초점

• 인터랙션 방식

- o dartpoint.ai: 정형화된 데이터와 리포트 제공, 사용자가 원하는 정보 선택 중심
- 우리 프로젝트: 자연어로 질문을 하면, 기업명/지표 추출 후 실제 데이터와 업종 비교를 통한 맞춤형 해석 및 판단 근거 제공

• 교육 및 사용자 경험

- dartpoint.ai: 분석 정보를 제공하지만, 교육적/해설적 측면은 상대적으로 약
 함
- 우리 프로젝트: 반복 질의-응답을 통해 사용자가 스스로 투자 판단 능력을 기를
 수 있도록 설계, 투자 학습자의 성장 지원

• 기술 응용 및 확장성

dartpoint.ai: 주로 DART 데이터 등 공시자료에 의존한 분석

우리 프로젝트: HuggingFace 기반 자연어 처리와 RAG 구조를 활용해 다양
 한 질문 유형에 대응하며, 맞춤형 투자 멘토링 확장이 용이

즉, dartpoint.ai가 **정적인 데이터 분석 서비스**라면, 우리 프로젝트는 **대화형 AI를 통한** 실시간 투자 판단 학습에 초점을 맞추어 차별화하고 있음.

▼ (강병진) 대화 예시

[예시 대화: 투자 판단력 훈련 챗봇 "INVestigator"]

사용자:

"삼성전자 PER이 18이면 괜찮은 건가요?"

챗봇 (INVestigator):

"현재 삼성전자의 PER은 18로 나타납니다. 같은 반도체 업종의 평균 PER은 약 15 정도인데, 이는 삼성전자가 업종 평균보다 다소 높은 평가를 받고 있음을 의미합니다.

- 해석1: 삼성전자가 미래 성장성에 대한 기대를 반영해 프리미엄이 붙은 상태일 수 있습니다.
- 해석2: 그러나 업종 내 경쟁 심화나 글로벌 시장 불확실성을 고려하면, 고평가로 볼 여지도 있습니다.

따라서 단순히 '괜찮다'고 단정짓기보다는, 현재 시장 상황과 기업의 미래 전략을 함께 검토할 필요가 있습니다."

사용자:

"그러면 이걸 어떻게 투자 판단에 반영해야 할까요?"

챗봇 (INVestigator):

"먼저, PER이 업종 평균보다 높은 이유를 파악하는 것이 중요합니다. 예를 들어:

- 성장 기대: 삼성전자의 경우, 미래 기술 투자나 글로벌 경쟁력 강화 계획이 반영되어 있을 수 있습니다.
- 리스크 요인: 반면, 반도체 시장의 경기 변동성, 원자재 가격 상승 등의 리스크도 존재합니다.

따라서, 투자 판단 시에는 PER 수치 외에도 기업의 재무 건전성, 성장 전략, 업계 트렌드 등을 종합적으로 고려해야 합니다.

추가적으로, 과거 실적 및 미래 전망 자료를 비교 분석하면 보다 신중한 결정을 내릴 수 있습니다."

사용자:

"그럼 지금 상황에서는 삼성전자가 고평가된 건가요?"

챗봇 (INVestigator):

"수치상으로는 삼성전자가 업종 평균보다 높은 PER을 보이고 있어, 일반적인 기준에서 는 고평가 신호로 해석될 수 있습니다. 하지만 고평가 여부는 단순 비교만으로 판단하기 어렵습니다.

- 긍정적 측면: 미래 성장 가능성과 기술 리더십이 반영된 결과일 수 있음
- 부정적 측면: 단기 시장 변동성과 외부 리스크가 반영된 결과일 수도 있음 따라서, 추가적인 재무 지표(예: EPS, PBR 등)와 시장 동향 분석을 함께 진행한 후 최종 판단하는 것이 좋습니다."

이 예시 대화는 단순히 정적인 정보를 제공하는 것이 아니라,

사용자가 스스로 투자 판단의 근거를 이해하고, 추가 분석 포인트를 학습할 수 있도록 돕는 구조임을 보여준다.

즉, 우리 챗봇은 **단순 정보 제공이 아닌, 투자 판단력을 향상시키는 멘토 역할**을 한다는 점이 차별화 포인트야.

▼ vs ChatGPT

ChatGPT: 범용 대화형 AI로 투자 관련 질문에 대해 일반적인 정보와 정의 제공에 그침 우리 프로젝트: "데이터 기반 해석 + 학습 피드백" 실시간 재무 데이터와 업종 비교를 기반으로, 투자 판단에 필요한 구체적 근거와 해석을 제공하며, 사용자가 스스로 판단력 을 향상할 수 있도록 피드백을 제공함

다트포인트: 정적인 공시 데이터와 재무제표 분석을 통해 자동 보고서를 생성하는 데 초 점

우리 프로젝트: 사용자가 자연어로 질문을 던지면, 해당 질문에 맞춰 내부 DB와 실시간 데이터를 연계해 맞춤형 분석과 업종 비교, 그리고 해석을 제공. 즉, 단순 정보 제공을 넘어 투자 판단력 향상을 위한 대화형 멘토 역할 수행

초보자가 가지는 어려움은 무엇인가? 자신에게 무엇이 부족한지, 무엇이 필요한지 잘 모른다는 점이다. 고수 투자자라면 어떤 정보를 어떤 소스에서 얻어야 할지, 어떤 관점을 택해야 할지 스스로 판단할 수 있다. 하지만 초보자는 정보의 홍수 속에서 막연하고 무섭다. INVestigator는 다양한 관점을 제공하며 투자자가 스스로 결정할 수 있는 판단의 근거를 제공하고 투자 감각을 길러주는 AI 멘토이다.

삼성전자 PER이 18이면 괜찮은 거야?

PER(주가수익비율)이 18이라면, 이는 삼성전자의 주가가 연간 순이익의 18배 수준에서 거래되고 있음을 의미합니다. 현재 삼성전자의 PER은 약 10.60배로, 과거에 비해 낮은 수준입니다. PER이 18이라면 현재보다 높아진 상태로, 이는 투자자들이 향후 이익 증가를 기대하고 주가를 높게 평가하고 있다는 해석이 가능합니다. 그러나 PER은 산업 평균, 기업의 성장 전망, 시장 상황 등 다양한 요소에 영향을 받으므로, 단순히 PER 수치만으로 투자 적정성을 판단하기는 어렵습니다. 따라서 PER 외에도 다른 재무 지표와 시장 동향을 종합적으로 고려하는 것이 바람직합니다. 별류라인: 투자자를 위한 종목발굴 박데이터 서비스







