금융 뉴스를 활용한
주식 포트폴리오 및 보고서 서비스

POSTECH IDS 강명구 장승현 홍현종

CONTENTS

01 Problem Definition

02 System Pipeline

03 Data collction and Preparation

04 Modeling

05 Result

1

Problem Definition

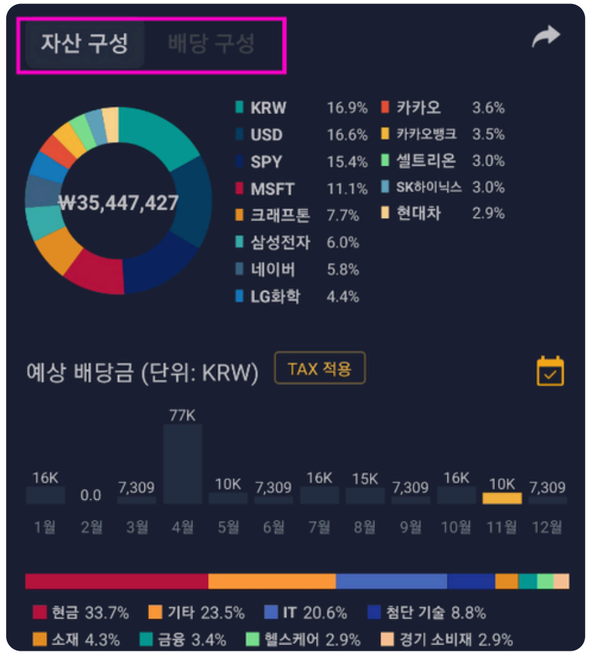
PROBLEM DEFINITION

주식 포트폴리오의 신뢰도 향상

Motivation

Portfolio Investment

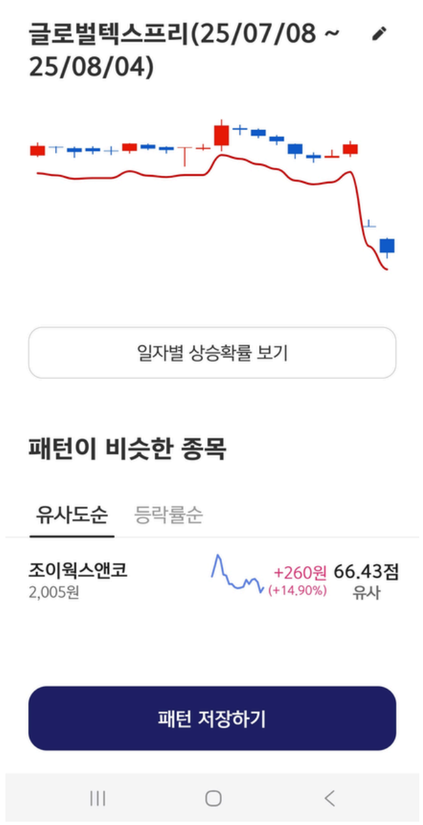
투자 위험을 분산하고 수익을 극대화하기 위해 여러 종목에 자산을 나누어 투자하는 주식 전략

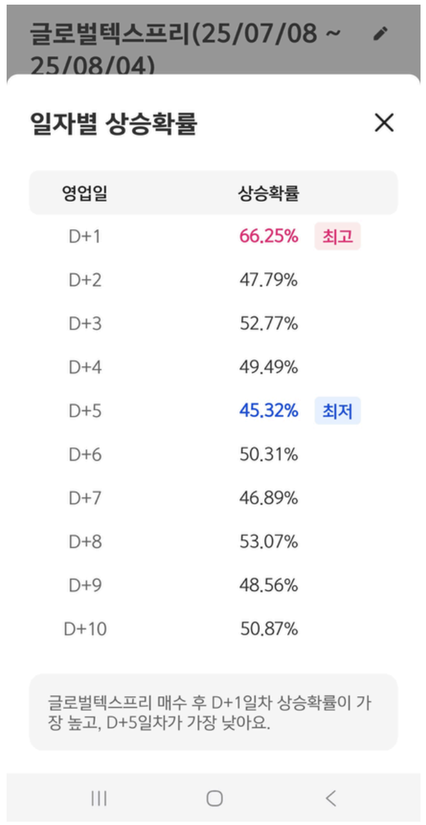


- 투자 성향 파악 : 단 / 장기 및 위험 감수 수준 정의
- 분산 투자 : 특정 종목에 자산을 집중하지 않음으로써 위험을 낮춤
- 자산 배분 : 국내외 및 적금 / 채권 등을 활용하여 자산을 배분

PROBLEM DEFINITION

주식 포트폴리오의 신뢰도 향상

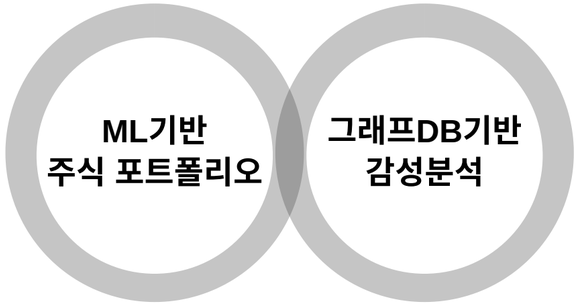




최근 금융 시장에는 **Pattern-based analysis, ML - based stock prediction**을 활용하여 자동 투자 포트폴리오 전략이 빠르게 확산

하지만 이런 접근 방식은 투자자에게 **설명력 부족과 신뢰도 저하 문제**를 동반하며, 특히 변동성이 큰 시장 환경에서 투자자가 결과를 그대로 **수용하기 어렵다는 한계**

Solution



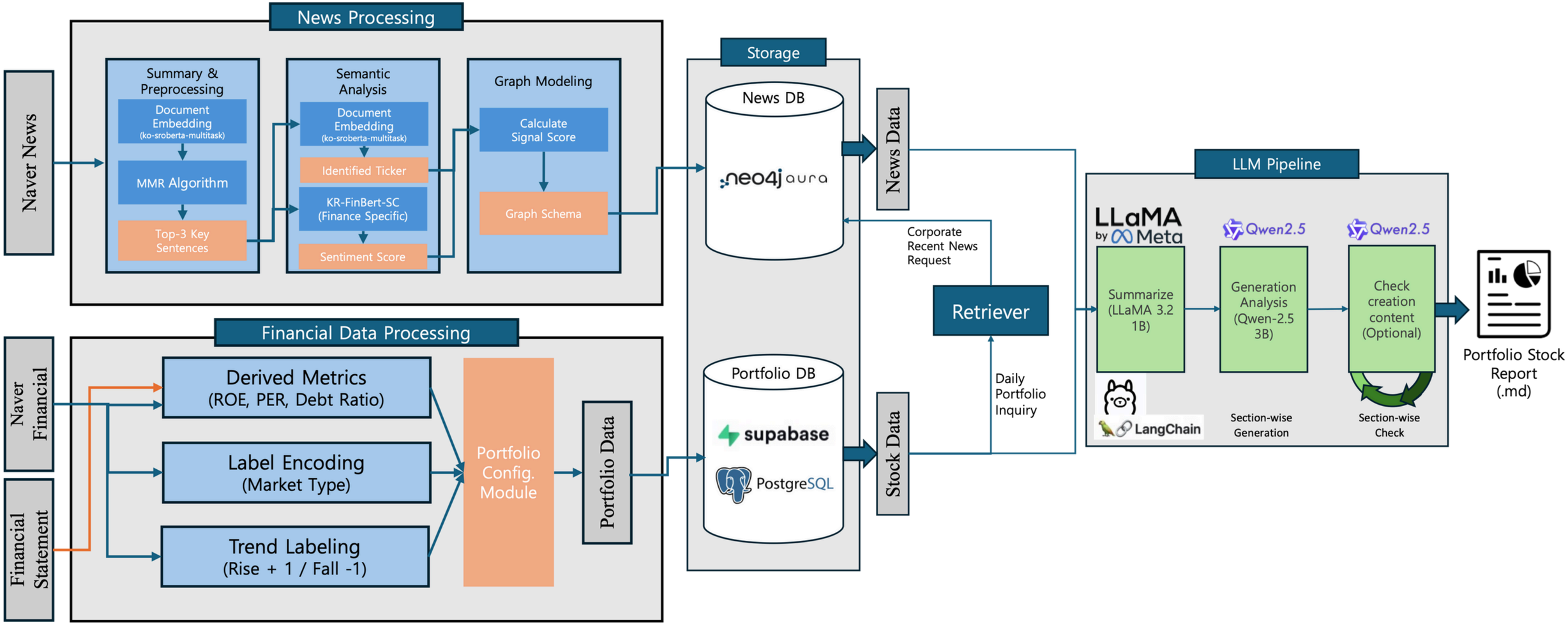
Key-Service

- 다양한 투자 전략 선택
- 전략적 기업 보고서

2

System Pipeline

OVERVIEW


News Processing
Naver News
Summary & Preprocessing
Document Embedding (ko-sroberta-multitask)
MMR Algorithm
Top-3 Key Sentences
Semantic Analysis
Document Embedding (ko-sroberta-multitask)
Identified Ticker
KR-FinBert-SC (Finance Specific)
Sentiment Score
Graph Modeling
Calculate Signal Score
Graph Schema
Financial Data Processing
Naver Financial
Financial Statement
Derived Metrics (ROE, PER, Debt Ratio)
Label Encoding (Market Type)
Trend Labeling (Rise + 1 / Fall -1)
Portfolio Config. Module
Portfolio Data
Storage
News DB
neo4j aura
Portfolio DB
supabase
PostgreSQL
News Data
Stock Data
Corporate Recent News Request
Daily Portfolio Inquiry
Retriever
LLM Pipeline
LLaMA by Meta
Qwen2.5
Qwen2.5
Summarize (LLaMA 3.2 1B)
Generation Analysis (Qwen-2.5 3B)
Check creation content (Optional)
LangChain
Section-wise Generation
Section-wise Check
Portfolio Stock Report (.md)

3

Data collction and Preparation

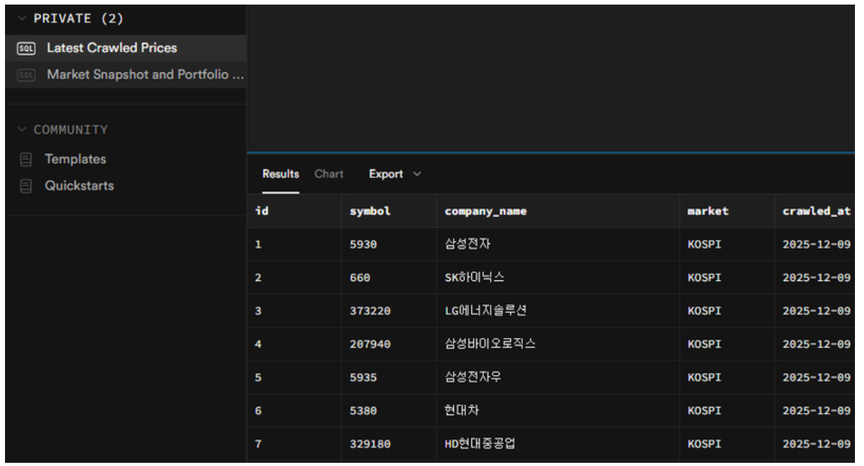
주식 데이터 수집 및 저장

Target : 네이버 증권의 코스피 / 코스닥 / 선물의 모든 카테고리를 수집

☐ 거래량	☐ 매수호가	☐ 거래대금(백만)	☐ 시가총액(억)	☐ 영업이익(억)	☐ PER(배)
☐ 시가	☐ 매도호가	☐ 전일거래량	☐ 자산총계(억)	☐ 영업이익증가율	☐ ROE(%)
☐ 고가	☐ 매수총잔량	☐ 외국인비율	☐ 부채총계(억)	☐ 당기순이익(억)	☐ ROA(%)
☐ 저가	☐ 매도총잔량	☐ 상장주식수(천주)	☐ 매출액(억)	☐ 주당순이익(원)	☐ PBR(배)
			☐ 매출액증가율	☐ 보통주배당금(원)	☐ 유보율(%)

- 매일 00:00 기준 전일 발생 데이터 전체를 수집

- 저장된 데이터는 crawled_data 이름으로 DB에 매일 갱신
  - 전일대비증감 열을 생성하여 +1 / -1로 처리
  - kospi / kosdaq / ETF 를 구분 후 LabelEncoder로 처리하여 투자 시 선택할 수 있게 적용
  - 자기자본이익 / 총자산이익 및 부채비율 / 자산회전율 등 분석에 필요한 재무 파생 변수 생성
    - 추가로 필요한 재무 정보는 TS-2000 회계 프로그램을 통해 데이터를 받아 모델의 feature로 사용



id	symbol	company_name	market	crawled_at
1	5930	삼성전자	KOSPI	2025-12-09
2	660	SK하이닉스	KOSPI	2025-12-09
3	373220	LG에너지솔루션	KOSPI	2025-12-09
4	207940	삼성바이오로직스	KOSPI	2025-12-09
5	5935	삼성전자우	KOSPI	2025-12-09
6	5380	현대차	KOSPI	2025-12-09
7	329180	HD현대중공업	KOSPI	2025-12-09

네이버 뉴스 데이터 수집 및 저장

Target : 네이버 뉴스의 경제, 세계, IT/과학 카테고리 수집

- 매일 00:00 기준 전일 발생 데이터 전체를 수집
- Composite Primary Key (복합 기본키) 사용
  - URL에서 언론사 ID와 기사 ID를 추출
  - (Journal_ID, Article_ID) 쌍을 Unique Key로 설정하여 DB 적재



https://n.news.naver.com/mnews/article/018/0006180564

Journal_ID

Article_ID

수집된 데이터는 아래의 형태로 PostgreSQL 저장

Journal_ID	Article_ID	Posted_Date	Title	Content	URL
언론사 ID	기사 ID	게재일	뉴스 제목	뉴스 본문	뉴스 URL

뉴스 데이터 전처리 1 - 추출형 요약

Extractive Summarization Strategy (추출 요약 전략)

- Objective: 뉴스 원문의 맥락(Context)을 유지하되, LLM 및 감성분석 모델의 연산 효율성을 위해 핵심 정보 압축
- Encoder Model: *jhgan/ko-sroberta-multitask* (Korean Sentence-BERT)

MMR (Maximal Marginal Relevance) Algorithm

- Logic: 단순 유사도 기반 추출 시 발생하는 정보 중복 문제를 해결하기 위해 관련성(Relevance)과 다양성(Diversity)을 동시에 고려

$$\lambda Sim_1(Relevane) - (1-\lambda)Sim_2(Diversity)$$

- Process
  - 문서 전체의 의미 벡터(Document Embedding) 산출
  - 문서 중심(Centroid)과 가깝지만, 이미 선택된 문장과는 거리가 먼 문장 Top-3 선별

뉴스 데이터 전처리 2 - 연관 기업 탐색 및 감성 분석

Morphological Entity Linking (형태소 기반 기업 식별)

- Methodology : *Kiwi* 형태소 분석기를 활용한 고유명사(NNP) 추출
  - 금융 도메인 특수성을 반영하기 위해 전체 상장 기업명(Ticker)을 사용자 사전에 등재
  - Effect: 'SK하이닉스', 'LG에너지솔루션' 등 복합 명사가 분절되지 않고 단일 토큰(Single Token)으로 인식

Financial Sentiment Analysis (금융 감성 분석)

- Model: *snunlp/KR-FinBert-SC* (금융 도메인 특화 Pre-trained BERT)
  - 감성 확신도(| Score |)가 0.1 미만인 모호한 기사는 Noise로 간주하여 그래프 구축에서 배제

```
MATCH (n:News)-[r:MENTIONS]->(c:Company {name: 'SK하이닉스'})
RETURN
    n.date AS 날짜,
    n.title AS 제목,
    // substring(속성, 시작위치, 길이)
    substring(n.summary, 0, 50) + '...' AS 요약_미리보기,
    r.sentiment AS 감성
ORDER BY n.date DESC, r.sentiment DESC
LIMIT 20
```

Table RAW

	날짜	제목	요약_미리보기	감성
1	2025-12-08	"[특징주 & 리포트] '하이닉스·삼성전자 나란히 강세' 등"	"김동원 KB증권 리서치본부장은 “내년부터 빅테크(대형 기술기업)의 주문형반도체(ASIC)가 ... "	0.9998372742884385
2	2025-12-08	"반도체 투톱 ‘30조 영업익’ 신기록 예고…칩 품귀에 호황 더 간다"	"SK하이닉스 16조원대 영업익 전망삼성전자도 역대 최대 분기익 눈앞AI發 전례 없는 메모리 ... "	0.9998339935445983

그래프DB 엣지 가중치 설계

Problem Statement (기존 한계점)

- Frequency Bias: 단순 언급 빈도(Count)가 높다고 해서 반드시 시장에 중요한 뉴스는 아님 (노이즈 가능성)
- Sentiment Intensity: 감성 점수만으로는 해당 이슈가 시장에서 얼마나 '주목(Attention)'받고 있는지 알 수 없음

Proposed Metric: Signal Score

- 뉴스의 주목도(Relevance)와 감성의 방향성(Sentiment)을 결합한 연결 강도 지표

$$Signal_{ij} = Sentiment_i \times \log(1 + Relevance_{ij})$$

$Sentimental_i$: FinBERT Score, $(-1 \leq S \leq 1)$

$Relevance_{ij}$ = 기업 hii에 대한 기사 j의 언급 가중치 합

  - 언급 가중치 : 제목(+3) , 본문(+1)

```
1 MATCH (n:News)-[r:MENTIONS]->(c:Company {name: 'SK하이닉스'})
2 RETURN
3     n.date AS 날짜,
4     n.title AS 제목,
5     // substring(속성, 시작위치, 길이)
6     substring(n.summary, 0, 50) + '...' AS 요약_미리보기,
7     r.weight AS Signal_Score
8 ORDER BY n.date DESC
9 LIMIT 20
```

Table RAW

2	2025-12-08	"[특징주 & 리포트] '하이닉스·삼성전자 나란히 강세' 등"	"김동원 KB증권 리서치본부장은 “내년부터 빅테크(대형 기술기업)의 주문형반도체(ASIC)가 ... "	0.6930343876917718
3	2025-12-08	"반도체 장비社 ASM, 화성 ‘혁신제조센터’ 가동…1억 달러 투입"	"행사에는 고영인 경기도 경제부지사와 조승문 화성특례시 제2부시장 등 지자체 관계자를 비롯해 ... "	0.4867057216611389

데이터 적재 및 그래프 스키마 설계

Property Graph Model Definition (속성 그래프 모델 정의)

- LPG (Labeled Property Graph) 모델을 채택하여, 뉴스 데이터와 기업 간의 관계를 구조화

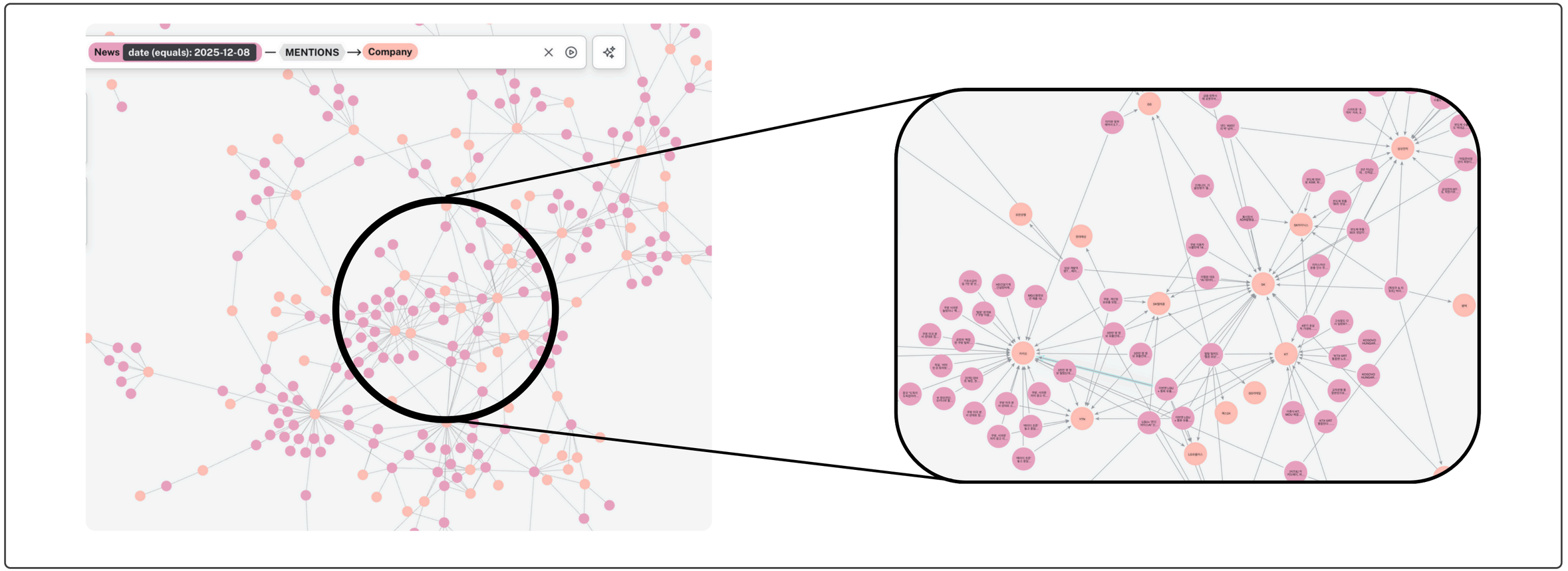
Node Specifications (노드 명세)

- **(N) News Node (:News)**
  - Identity : (journal_id, article_id) (Composite PK)
  - Attributes :
    - summary: SBERT 기반 3문장 요약본 (For LLM Reporting)
    - date: 기사 발행일 (YYYY-MM-DD)
- **(C) Company Node (:Company)**
  - Identity: name (Unique Key, Ticker Name)

Relationship Specifications (관계 명세)

- **(R) Mentions Edge [:MENTIONS]**
  - Direction: (:News) -> (:Company)
  - Attributes :
    - summary: SBERT 기반 3문장 요약본 (For LLM Reporting)
    - date: 기사 발행일 (YYYY-MM-DD)

데이터 적재 및 그래프 스키마 설계



4

Modeling

AI Selection

Feature Engineering

- Price - based
  - Moving Average
  - Return
  - Price - Momemtum
- Volume - based
  - Change - Rate
  - OBV(On-Balance - Volume) - 추세의 힘
  - Turnover Value - 거래 대금
- Hybrid - Feature
  - RSI
  - MACD
  - Bollinger Bands
- Structural Feature
  - Candlestick - Pattern
  - High - Low Range

Model Selection

- Model Selection - XGBOOST
- train / test split : 과거 80% 시점 / 최근 20% 시점
- 유저가 설정한 threshold 기간 및 수익률 이상으로 상승 / 하락을 예측 후 확률 출력

Optimization & Validation

Portfolio Optimization

$$Sharpe = \frac{R_q - R_f}{\sigma_p}$$

R_p 포트폴리오 기대 수익률

R_f 무위험 이자율

σ_p 포트폴리오의 변동성 (표준편차)

- 몬테카를로 시뮬레이션을 통한 수만 개의 포트폴리오 조합 생성
- Sharpe기반 목적 함수를 최대화하는 비중으로 채택
- 각 종목 간 공분산 행렬을 계산하여 종목 A가 상승할 때 종목 B의 상관 관계 파악 및 Efficient Frontier상 에서의 Optimal Weights 출력

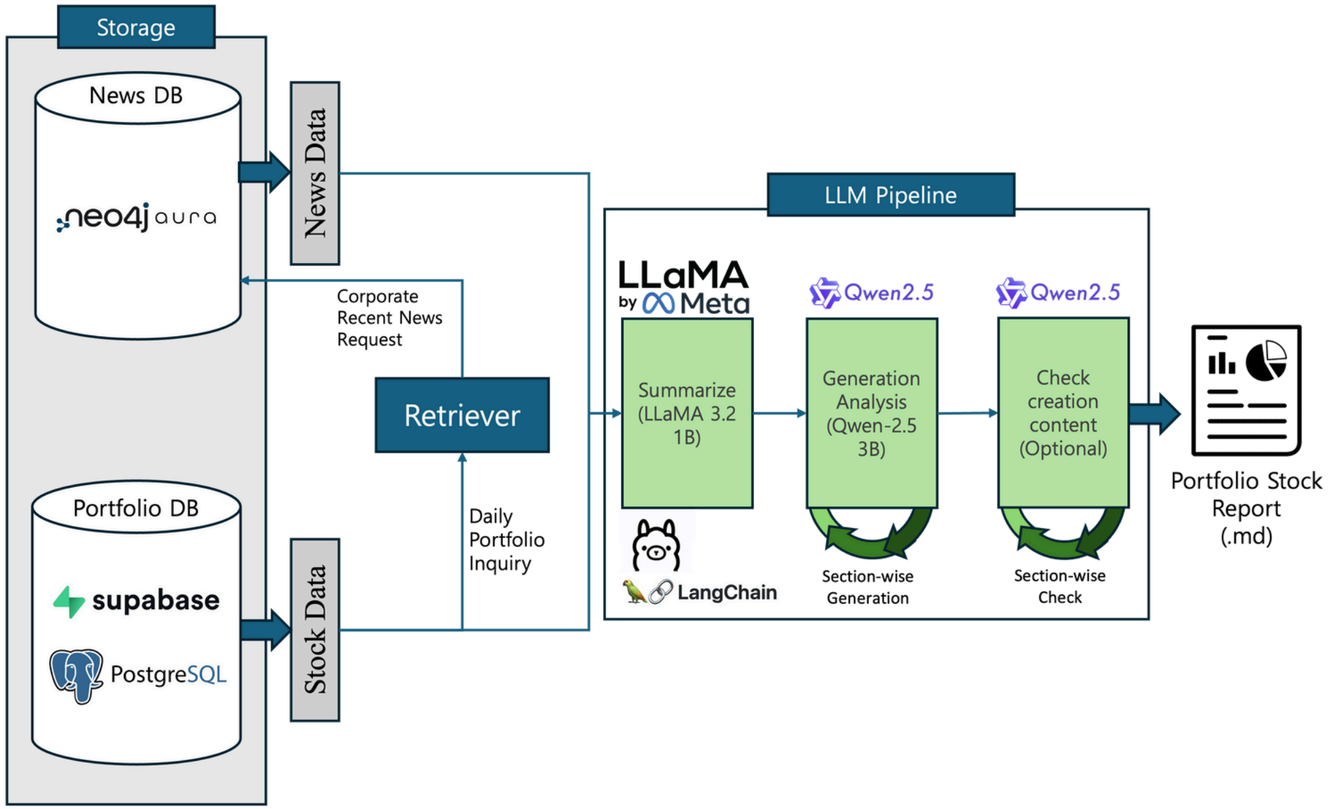
Strategic Backtesting

- RSI / BB / MACD / SO 등 4가지 관점에서 백테스팅 진행
- CAGR(연평균 성장률) , MDD(최대 낙폭) , Win Rate(승률)의 세 가지를 중심으로 백테스트 결과를 평가

Entry Points Suggestion

- 과거 MA의 교차점 분석을 통한 Sell/Buy signal을 통해 매수 시점 추천

LLM파트



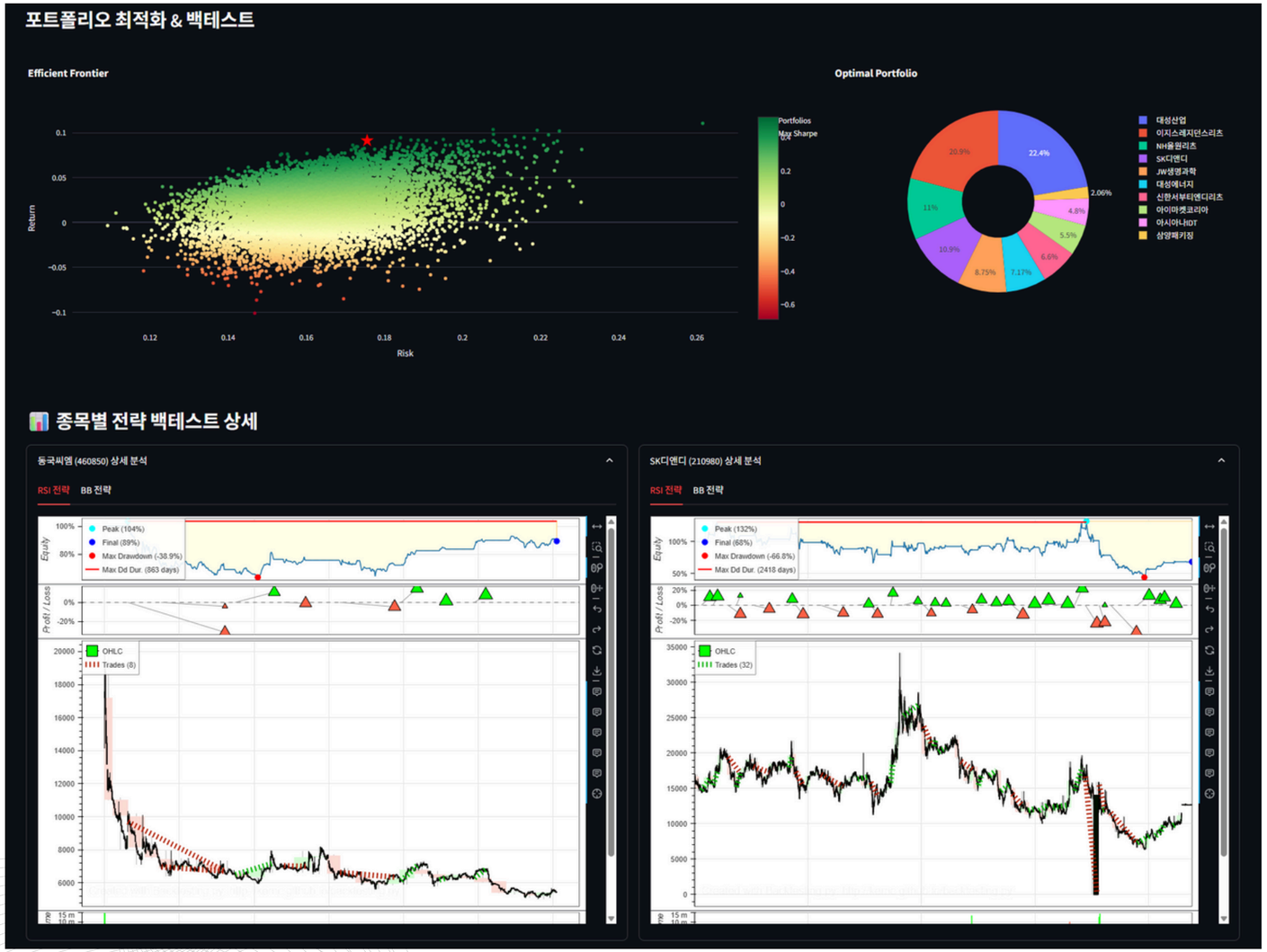
- **LangChain, Ollama 이용**
  - In-Context Learning 활용 - Report 텍스트 프롬프트로 제공
- **Model Specialization**
  - Summarization: LLaMA-3.2 (1B) 모델 사용하여 종목별 뉴스 데이터 요약
  - Generation & Validation: Qwen-2.5 (3B)를 사용하여 섹션별 본문 작성 및 형식 검증
- **Iterative Generation Pipeline (LangChain)**
  - 단일 패스 생성이 아닌, Section-wise Loop 구조를 채택하여 긴 리포트의 내용 경량 모델로 작성
  - 생성 후 포맷 검증(Check) 단계 - 최종 결과물(.md)의 안정성 확보

5

Result

DEMO

user... t...	name text	sou... t...	created_at timestamptz
demo_user	Auto-20251209	Streamlit App v	2025-12-09 10:52:00.998607+0
demo_user	Auto-Port-20251209-2002	Streamlit App v	2025-12-09 11:02:23.760674+0(
demo_user	Auto-Port-20251209-2004	Streamlit App v	2025-12-09 11:04:51.508864+0(
demo_user	Auto-Port-20251209-2006	Streamlit App v	2025-12-09 11:06:18.909004+0(
demo_user	Auto-Port-20251209-2009	Streamlit App v	2025-12-09 11:09:32.008279+0(
demo_user	Auto-Port-20251209-2009	Streamlit App v	2025-12-09 11:09:41.469332+0(
demo_user	Auto-Port-20251209-2009	Streamlit App v	2025-12-09 11:09:52.926194+0(
demo_user	Auto-Port-20251209-2016	Streamlit App v	2025-12-09 11:16:25.258303+00
demo_user	Auto-Port-20251209-2016	Streamlit App v	2025-12-09 11:16:31.852916+00
demo_user	Auto-Port-20251209-2016	Streamlit App v	2025-12-09 11:16:52.56713+00
demo_user	Auto-Port-20251209-2022	Streamlit App v	2025-12-09 11:22:13.458779+00
demo_user	Auto-Port-20251209-2024	Streamlit App v	2025-12-09 11:24:31.551143+00
demo_user	Auto-Port-20251209-2027	Streamlit App v	2025-12-09 11:27:20.211445+00
demo_user	Auto-Port-20251210-1310	Streamlit App v	2025-12-10 04:10:49.04058+00
demo_user	Auto-Port-20251210-1311	Streamlit App v	2025-12-10 04:11:37.11695+00
demo_user	Auto-Port-20251210-1318	Streamlit App v	2025-12-10 04:18:33.329258+0(
demo_user	Auto-Port-20251210-1320	Streamlit App v	2025-12-10 04:20:37.886954+0(
demo_user	Auto-Port-20251210-1320	Streamlit App v	2025-12-10 04:20:59.491634+0(
demo_user	Auto-Port-20251210-1321	Streamlit App v	2025-12-10 04:21:17.302643+00
demo_user	Auto-Port-20251210-1321	Streamlit App v	2025-12-10 04:21:56.021108+00
demo_user	Auto-Port-20251210-1322	Streamlit App v	2025-12-10 04:22:23.882946+0(
demo_user	Auto-Port-20251210-1323	Streamlit App v	2025-12-10 04:23:18.366631+0(



DEMO


국내주식 AI 하이브리드 포트폴리오
Supabase DB + AI + 최적 MA 전략 자동 탐색
1 AI 종목선정
2 포트폴리오 & 백테스트
3 최적 매수 타이밍
포트폴리오 최적화 & 백테스트
Efficient Frontier
Return
Risk
Portfolios
Max Sharpe
Optimal Portfolio
케이씨텍
프레스티지바이오파마
해성디에스
PI첨단소재
덴티움
39.9%
21.4%
종목별 전략 백테스트 상세
덴티움 (145720) 상세 분석
RSI 전략
BB 전략
Peak (217%)
Final (79%)
케이씨텍 (281820) 상세 분석
PI첨단소재 (178920) 상세 분석
일진하이솔루스 (271940) 상세 분석

📊 [Daily Portfolio] 시장 주도주 점검 및 대응 전략

> (서론: 2025년 12월 10일 기준, 오늘 분석된 종목들의 섹터(반도체) 분위기와 핵심 이슈를 종합하여 5-7문장으로 상세히 요약. 시장 전반의 흐름과 주요 변동 요인을 포함)

1. 포트폴리오 요약

종목명	핵심 키워드	투자 의견	주요 이슈
케이씨텍	AI, 반도체	매수	삼성 '평택 5공장' 시동 및 마키나락스의 AI 에이전트 개발
해성디에스	반도체	관망	3분기 영업익 발표 및 DDR5용 기판 공급 확대
프레스티지바이오파마	바이오, K바이오	매수	CTHRC1 개발 및 미국 의약품 수출 전략
덴티움	치과용 의료기기	관망	주요 연구개발 프로젝트 중단

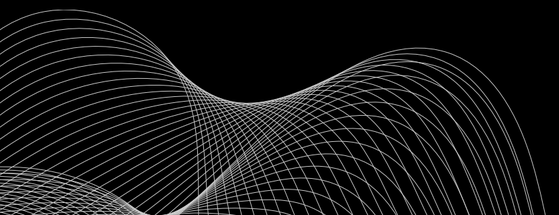
2. 종목별 심층 분석 및 전망

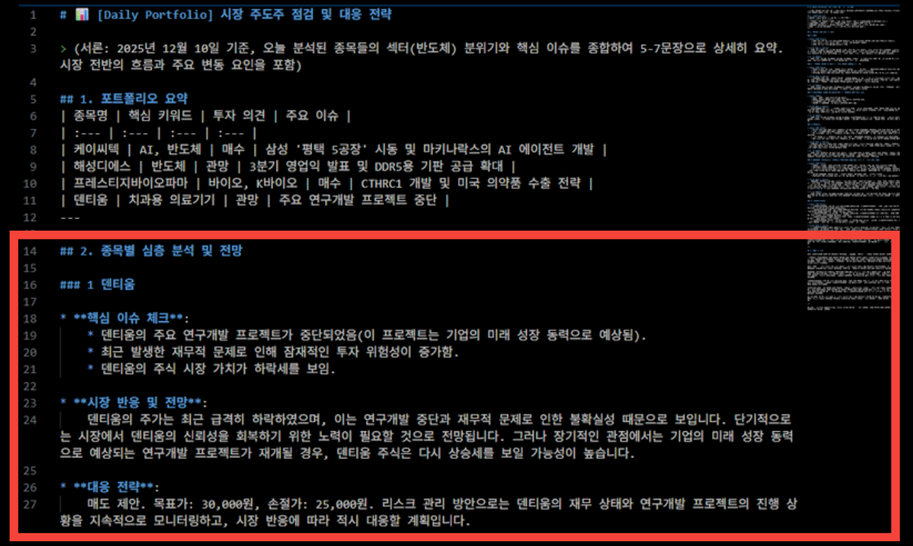
1 덴티움

- **핵심 이슈 체크:**
  - 덴티움의 주요 연구개발 프로젝트가 중단되었음(이 프로젝트는 기업의 미래 성장 동력으로 예상됨).
  - 최근 발생한 재무적 문제로 인해 잠재적인 투자 위험성이 증가함.
  - 덴티움의 주식 시장 가치가 하락세를 보임.
- **시장 반응 및 전망:** 덴티움의 주가는 최근 급격히 하락하였으며, 이는 연구개발 중단과 재무적 문제로 인한 불확실성 때문으로 보입니다. 단기적으로는 시장에서 덴티움의 신뢰성을 회복하기 위한 노력이 필요할 것으로 전망됩니다. 그러나 장기적인 관점에서는 기업의 미래 성장 동력으로 예상되는 연구개발 프로젝트가 재개될 경우, 덴티움 주식은 다시 상승세를 보일 가능성이 높습니다.
- **대응 전략:** 매도 제안. 목표가: 30,000원, 손절가: 25,000원. 리스크 관리 방안으로는 덴티움의 재무 상태와 연구개발 프로젝트의 진행 상황을 지속적으로 모니터링하고, 시장 반응에 따라 적시 대응할 계획입니다.

2 케이씨텍: KEC의 TC 본더 4 선보여 반도체 신뢰성 개선 기대

- **핵심 이슈 체크:**
  - KEC의 TC 본더 4는 온도 변화에도 목표한 성능을 유지할 수 있는 고대역폭메모리(HBM4)용 열압착(TC) 본더를 선보였다.





```
# 📊 [Daily Portfolio] 시장 주도주 점검 및 대응 전략

> (서론: 2025년 12월 10일 기준, 오늘 분석된 종목들의 섹터(반도체) 분위기와 핵심 이슈를 종합하여 5-7문장으로 상세히 요약. 시장 전반의 흐름과 주요 변동 요인을 포함)

## 1. 포트폴리오 요약
| 종목명 | 핵심 키워드 | 투자 의견 | 주요 이슈 |
| :--- | :--- | :--- | :--- |
| 케이씨텍 | AI, 반도체 | 매수 | 삼성 '평택 5공장' 시동 및 마키나락스의 AI 에이전트 개발 |
| 해성디에스 | 반도체 | 관망 | 3분기 영업익 발표 및 DDR5용 기판 공급 확대 |
| 프레스티지바이오파마 | 바이오, K바이오 | 매수 | CTHRC1 개발 및 미국 의약품 수출 전략 |
| 덴티움 | 치과용 의료기기 | 관망 | 주요 연구개발 프로젝트 중단 |
---

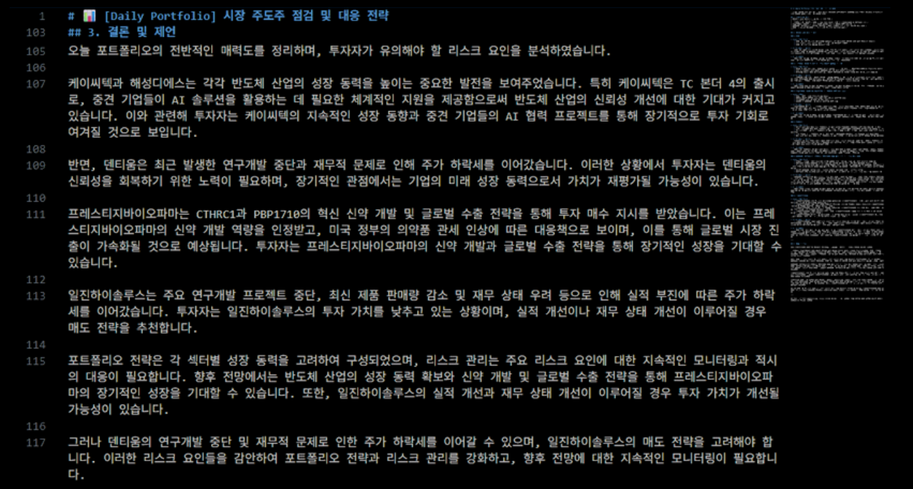
## 2. 종목별 심층 분석 및 전망

### 1 덴티움

* **핵심 이슈 체크**:
    * 덴티움의 주요 연구개발 프로젝트가 중단되었음(이 프로젝트는 기업의 미래 성장 동력으로 예상됨).
    * 최근 발생한 재무적 문제로 인해 잠재적인 투자 위험성이 증가함.
    * 덴티움의 주식 시장 가치가 하락세를 보임.

* **시장 반응 및 전망**:
    덴티움의 주가는 최근 급격히 하락하였으며, 이는 연구개발 중단과 재무적 문제로 인한 불확실성 때문으로 보입니다. 단기적으로는 시장에서 덴티움의 신뢰성을 회복하기 위한 노력이 필요할 것으로 전망됩니다. 그러나 장기적인 관점에서는 기업의 미래 성장 동력으로 예상되는 연구개발 프로젝트가 재개될 경우, 덴티움 주식은 다시 상승세를 보일 가능성이 높습니다.

* **대응 전략**:
    매도 제안. 목표가: 30,000원, 손절가: 25,000원. 리스크 관리 방안으로는 덴티움의 재무 상태와 연구개발 프로젝트의 진행 상황을 지속적으로 모니터링하고, 시장 반응에 따라 적시 대응할 계획입니다.
```



```
# 📊 [Daily Portfolio] 시장 주도주 점검 및 대응 전략
## 3. 결론 및 제언
오늘 포트폴리오의 전반적인 매력도를 정리하며, 투자자가 유의해야 할 리스크 요인을 분석하였습니다.

케이씨텍과 해성디에스는 각각 반도체 산업의 성장 동력을 높이는 중요한 발전을 보여주었습니다. 특히 케이씨텍은 TC 본더 4의 출시로, 중견 기업들이 AI 솔루션을 활용하는 데 필요한 체계적인 지원을 제공함으로써 반도체 산업의 신뢰성 개선에 대한 기대가 커지고 있습니다. 이와 관련해 투자자는 케이씨텍의 지속적인 성장 동향과 중견 기업들의 AI 협력 프로젝트를 통해 장기적으로 투자 기회로 여겨질 것으로 보입니다.

반면, 덴티움은 최근 발생한 연구개발 중단과 재무적 문제로 인해 주가 하락세를 이어갔습니다. 이러한 상황에서 투자자는 덴티움의 신뢰성을 회복하기 위한 노력이 필요하며, 장기적인 관점에서는 기업의 미래 성장 동력으로서 가치가 재평가될 가능성이 있습니다.

프레스티지바이오파마는 CTHRC1과 PBP1710의 혁신 신약 개발 및 글로벌 수출 전략을 통해 투자 매수 지시를 받았습니다. 이는 프레스티지바이오파마의 신약 개발 역량을 인정받고, 미국 정부의 의약품 관세 인상에 따른 대응책으로 보이며, 이를 통해 글로벌 시장 진출이 가속화될 것으로 예상됩니다. 투자자는 프레스티지바이오파마의 신약 개발과 글로벌 수출 전략을 통해 장기적인 성장을 기대할 수 있습니다.

일진하이솔루스는 주요 연구개발 프로젝트 중단, 최신 제품 판매량 감소 및 재무 상태 우려 등으로 인해 실적 부진에 따른 주가 하락세를 이어갔습니다. 투자자는 일진하이솔루스의 투자 가치를 낮추고 있는 상황이며, 실적 개선이나 재무 상태 개선이 이루어질 경우 매도 전략을 추천합니다.

포트폴리오 전략은 각 섹터별 성장 동력을 고려하여 구성되었으며, 리스크 관리는 주요 리스크 요인에 대한 지속적인 모니터링과 적시의 대응이 필요합니다. 향후 전망에서는 반도체 산업의 성장 동력 확보와 신약 개발 및 글로벌 수출 전략을 통해 프레스티지바이오파마의 장기적인 성장을 기대할 수 있습니다. 또한, 일진하이솔루스의 실적 개선과 재무 상태 개선이 이루어질 경우 투자 가치가 개선될 가능성이 있습니다.

그러나 덴티움의 연구개발 중단 및 재무적 문제로 인한 주가 하락세를 이어갈 수 있으며, 일진하이솔루스의 매도 전략을 고려해야 합니다. 이러한 리스크 요인들을 감안하여 포트폴리오 전략과 리스크 관리를 강화하고, 향후 전망에 대한 지속적인 모니터링이 필요합니다.
```