



주식 투자 판단을 위한
과거 유사 뉴스 제공 서비스

1 서비스 개요

2 서비스 시연

3 기능 개발

4 아키텍처

5 트러블 슈팅

6 디벨롭

1 서비스 개요

트럼프 "약값 최대 80% 내릴 것"...

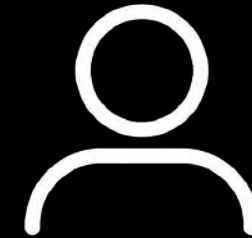
2025.05.21



도널드 트럼프 미국 대통령이 집권 1기에 이어 또다시 미국 내 약값을 인하하는 내용의 행정명령에 12일 서명했다.

트럼프 "약값 최대 80% 내릴 것"...

2025.05.21



일반인

"미국 환자의 약값 부담이 적어지겠구나"

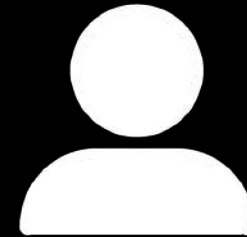
도널드 트럼프 미국 대통령이 집권 1기에 이어 또다시 미국 내 약값을 인하하는 내용의 행정명령에 12일 서명했다.

트럼프 "약값 최대 80% 내릴 것"...

2025.05.21



도널드 트럼프 미국 대통령이 집권 1기에 이어 또다시 미국 내 약값을 인하하는 내용의 행정명령에 12일 서명했다.



주식 투자자

“한국 제약 회사 주가에 영향이 있을까?”

“집권 1기 당시에는 주가가 어땠을까?”

기업 이슈에 따른 주가 변동 발생

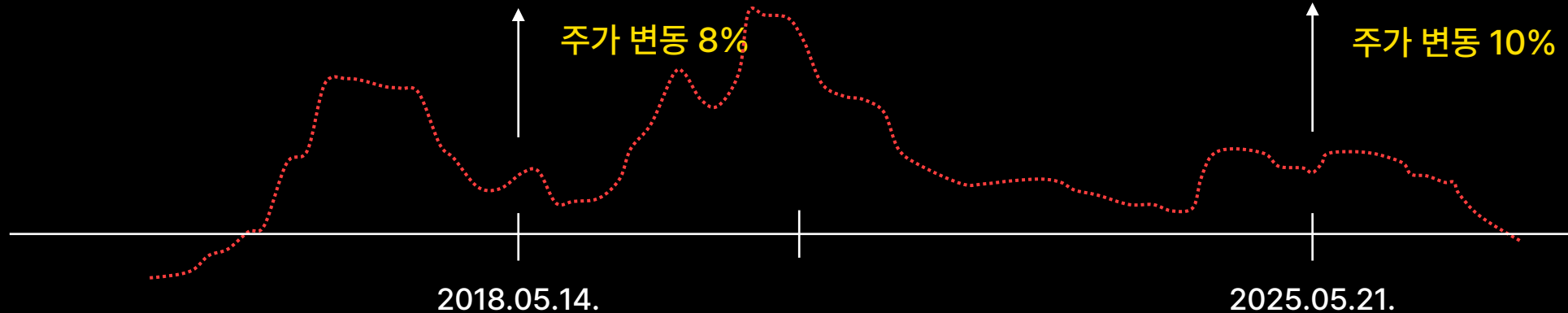
셀트리온 주가 차트

셀트리온, 美 트럼프 정부
약값 인하정책에 '방긋'

트럼프 대통령의 약값 인하
정책으로 그간 보수적이던 미국
바이오시밀러 허가 문턱이
낮아질 것으로 기대되는 만큼,
국내 기업 중에서는 셀트리온이
가장 큰 수혜기업으로 떠오르고
있다.

트럼프 '약값
행정명령'에도
...K-바이오는 반등

한국의 바이오시밀러(바이오
복제 약) 기업들에게는 이번 조치가
기회가 될 것으로 전망된다.
오리지널 의약품보다 값이 싼
바이오시밀러 처방이 확대돼
반사이익이 예상된다는 것이다.



1 서비스 개요

주식 투자 판단 도움이 되는 인사이트 제공

셀트리온, 美 트럼프 정부
약값 인하정책에 '방긋'

트럼프 대통령의 약값 인하
정책으로 그간 보수적이던 미국
바이오시밀러 허가 문턱이 낮아질
것으로 기대되는 만큼, 국내 기업
중에서는 셀트리온이 가장 큰
수혜기업으로 떠오르고 있다.

뉴스 중요도: 78점



트럼프 '약값
행정명령'에도
...K-바이오는 반등

한국의 바이오시밀러(바이오 복제
약) 기업들에는 이번 조치가
기회가 될 것으로 전망된다.
오리지널 의약품보다 값이 싼
바이오시밀러 처방이 확대돼
반사이익이 예상된다는 것이다.

유사도: 0.84

미국 약값 인하 기대에
셀트리온, 바이오시밀러
수혜주로 부각

미국 정부의 약가 규제 완화로
바이오시밀러 진입 장벽이
낮아지면서 셀트리온이 최대
수혜주로 주목받고 있다.

유사도: 0.78

美 약가 인하 정책 혼풍...
셀트리온, 글로벌 시장
공략 청신호

미국 약가 인하 정책으로 규제
장벽이 완화되면서 셀트리온의
바이오시밀러 해외 시장 진출과
매출 확대가 더욱 가속화될
전망이다.

유사도: 0.65

주가에 영향을 미치는
뉴스 선별

과거 유사 사건 뉴스 탐색

2 서비스 시연

2 서비스 시연

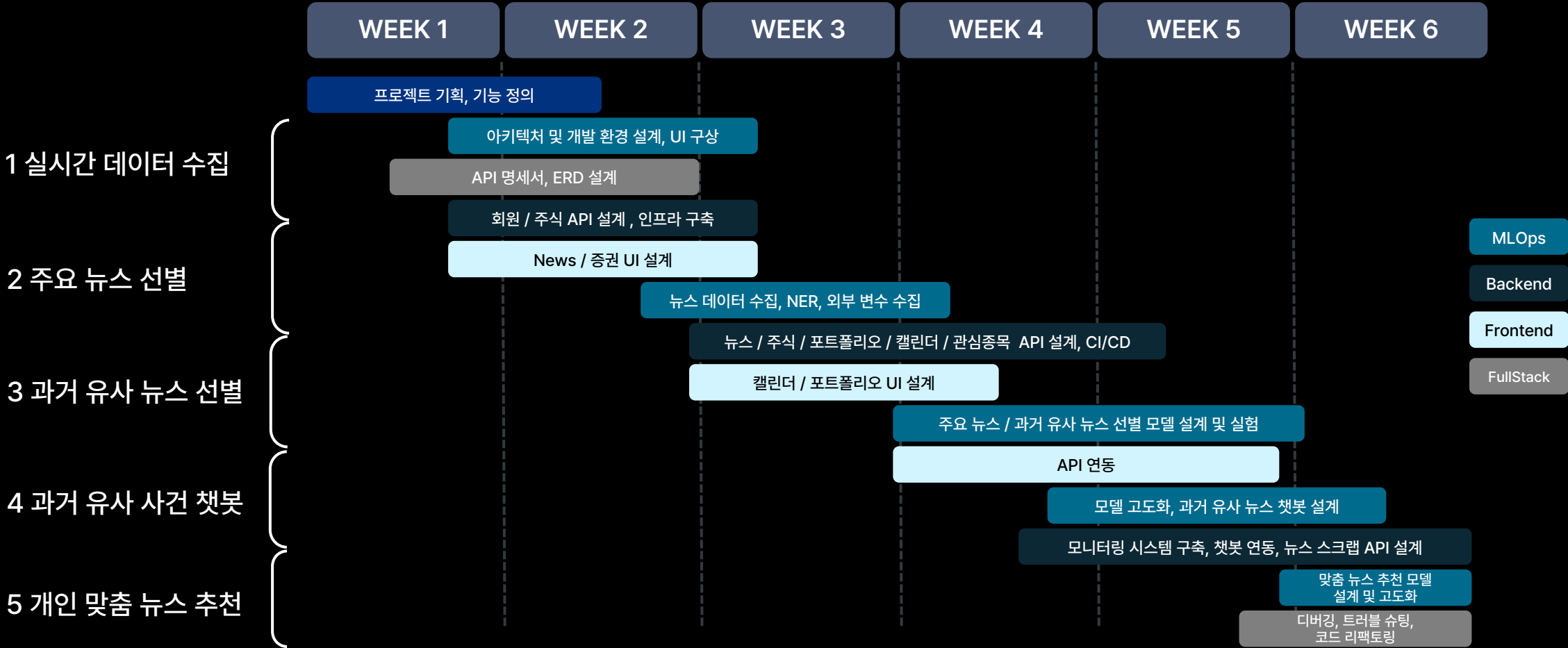


FROM NEWS TO INSIGHT

news-toss.vercel.app

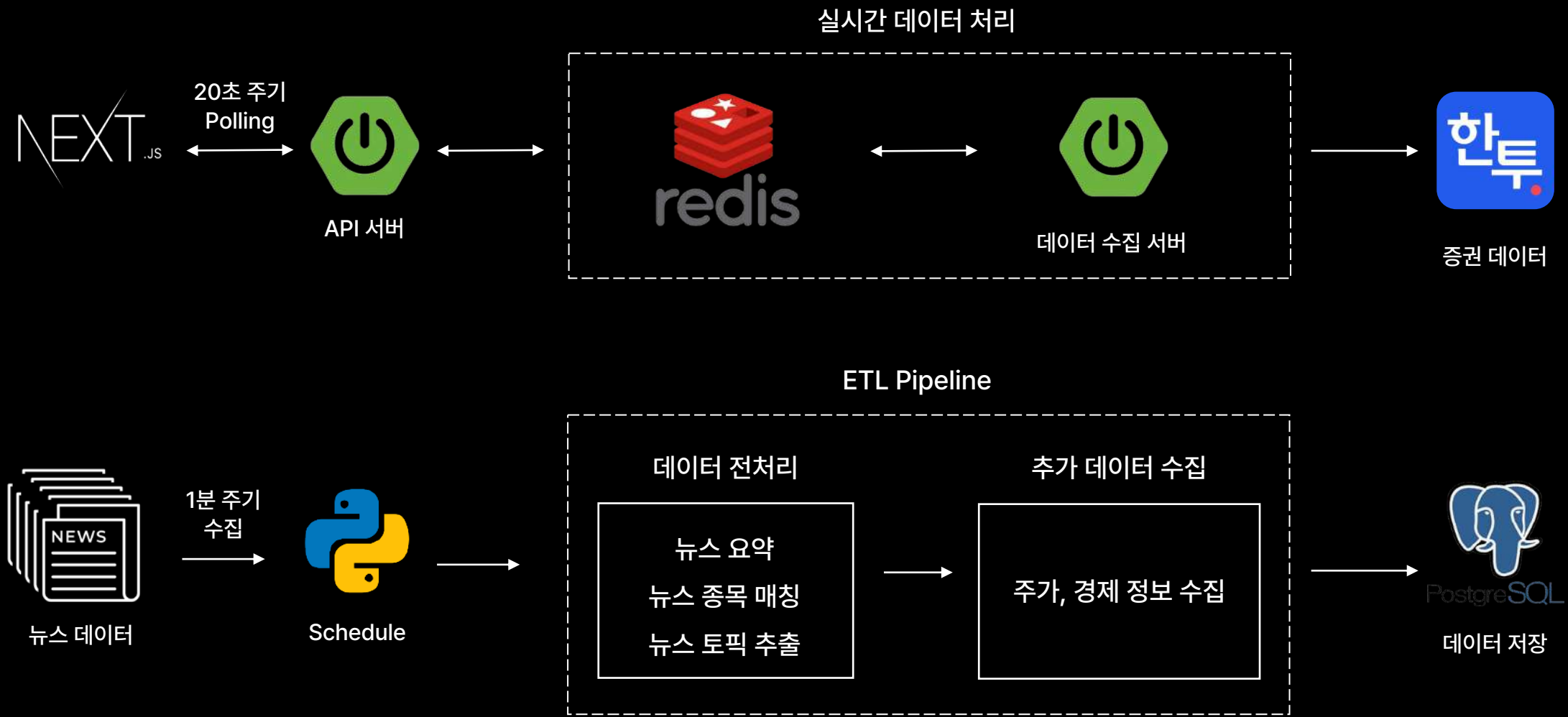
3 기능 개발

3-0 개발 일정

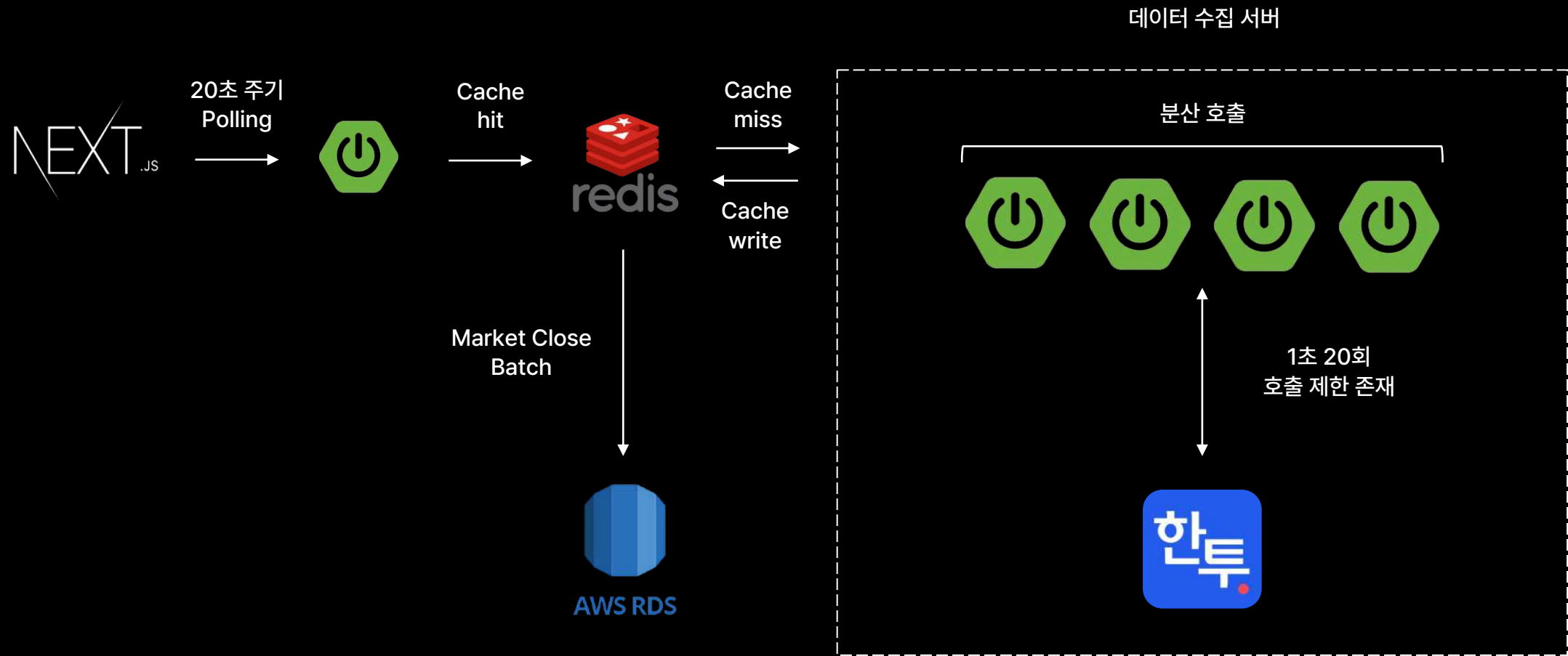


3-1 실시간 데이터 수집

3-1 실시간 데이터 수집



3-1 실시간 데이터 수집



3-1 실시간 데이터 수집

엔비디아가 역대 최고가를 경신하자
SK하이닉스(000660) 주가도 덩달아 뛰었다.
SK하이닉스는 엔비디아에 인공지능(AI) 반도체 핵심 구성품인 **고대역폭메모리(HBM)**를 공급하고 있다. 증권가에서는 SK하이닉스 목표가를 연달아 높여 잡고 있다.

토픽 모델링
LDA

- 대기업
사업 전략
- 반도체 / AI
- 기술 개발

데이터 전처리

→ 주식 종목 매칭
KDF/KDF-BERT-NER

→ 사건 중심 요약
Digit82/KoBART-Summarization

SK하이닉스(000660)

발행일	종목	종가	거래량	외국인 순매수	기관 순매수	개인 순매수	기준 금리	국채 10년 수익률	환율
24-06-22	SK하이닉스	198.0132	534587	432842	324438	44345365	1.0	2.0	1368.0
24-06-21	SK하이닉스	198.0115	475654	658844	246948	36546478	1.0	2.0	1365.0
24-06-20	SK하이닉스	197.9877	698976	512235	347895	67898712	1.0	2.0	1366.0
24-06-19	SK하이닉스	198.0173	365479	448976	125684	57897135	1.0	2.0	1369.0

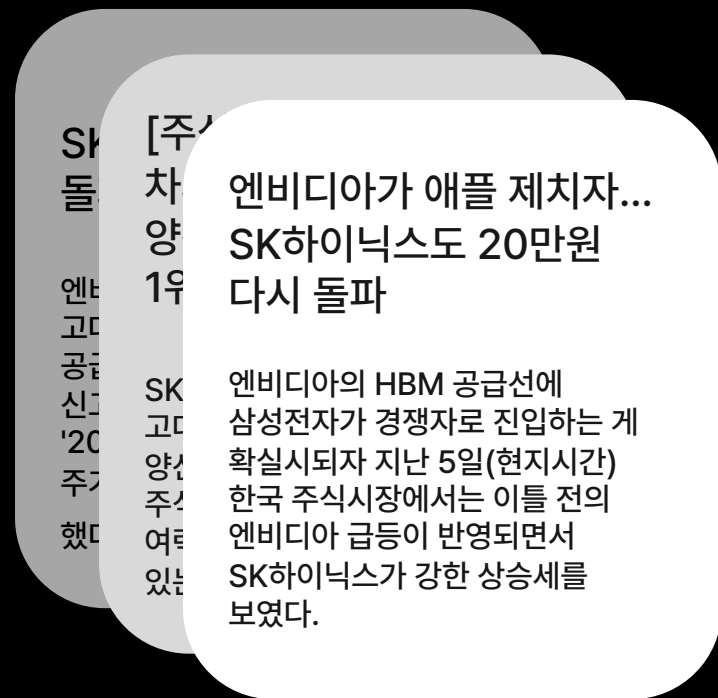
외부 시장 변수
(D-5 ~ D+5 종가, 거래량, 환율, 금리, 국채 등)

추가 데이터 수집



3-2 주요 뉴스 선별

3-2 주요 뉴스 선별



하루 200건의 수집 뉴스

어떤 뉴스를
주요 뉴스로 제공해야 할까?

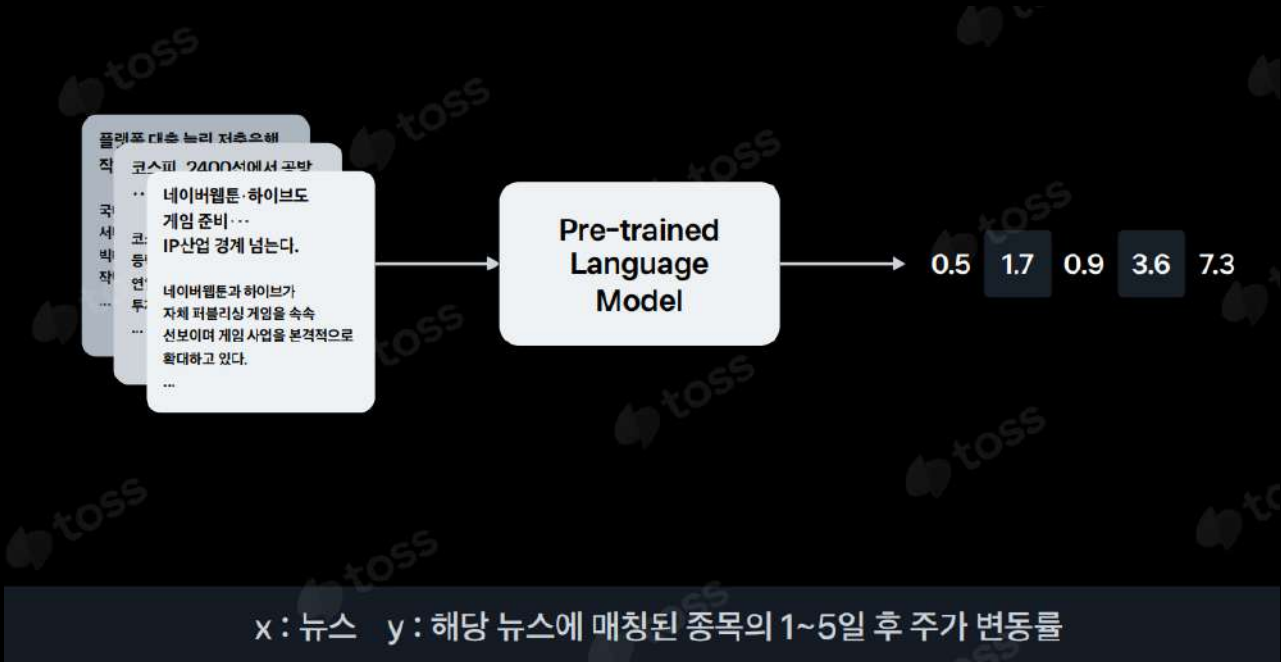
3-2 주요 뉴스 선별



Ref. Toss SLASH 23

머신러닝으로 더 똑똑하게 증권 뉴스 제공하기

3-2 주요 뉴스 선별



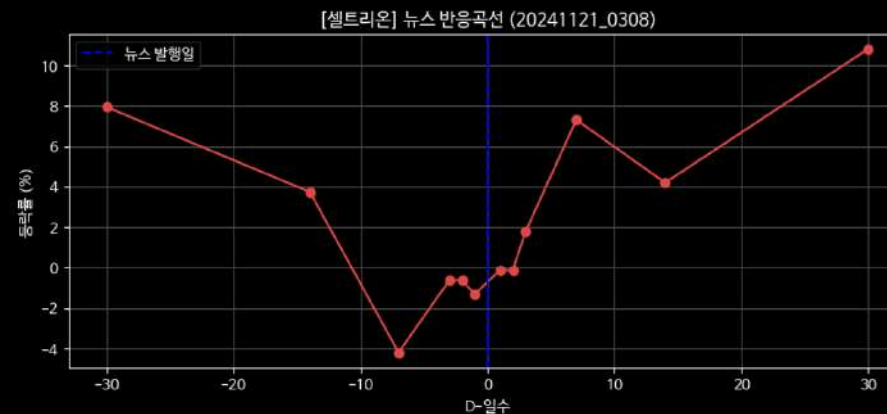
제목	abs (impact score)
아모레퍼시픽 목표가 19% 상향, 중국수요 빠른 개선-KTB證	3.5
봄 맞은 아모레퍼시픽, 회복 넘은 체질 개선으로 기대감↑	1.71
[클릭 e종목]"中 화장품 소비 폭발"...아모레퍼시픽 목표가 '29만'	1.08
KTB투자증권 "아모레퍼시픽, 중국 수요 개선 예상보다 빨라"	0.98
11번가 '아모레파티' 연다 ... 라방에서 득템하자	0.54
나이는숫자~관리는 필수~아모레 파티~	0.32
ESG가 기업이익 줄이는 '굴레' 안되게...성과지표 표준화 서둘러야	0.14
11번가, 아모레퍼시픽과 단독 행사...20% 할인 쿠폰	0.07

뉴스 중요도 계산

3-2 주요 뉴스 선별

셀트리온, 내년 신규
시밀러 품목 확대...실적
성장 전망

셀트리온은 2025년 최대
5개의 신규 시밀러 품목을
출시할 예정으로 현재 6개
품목에서 11개까지 시밀러
품목이 늘어날 전망이다

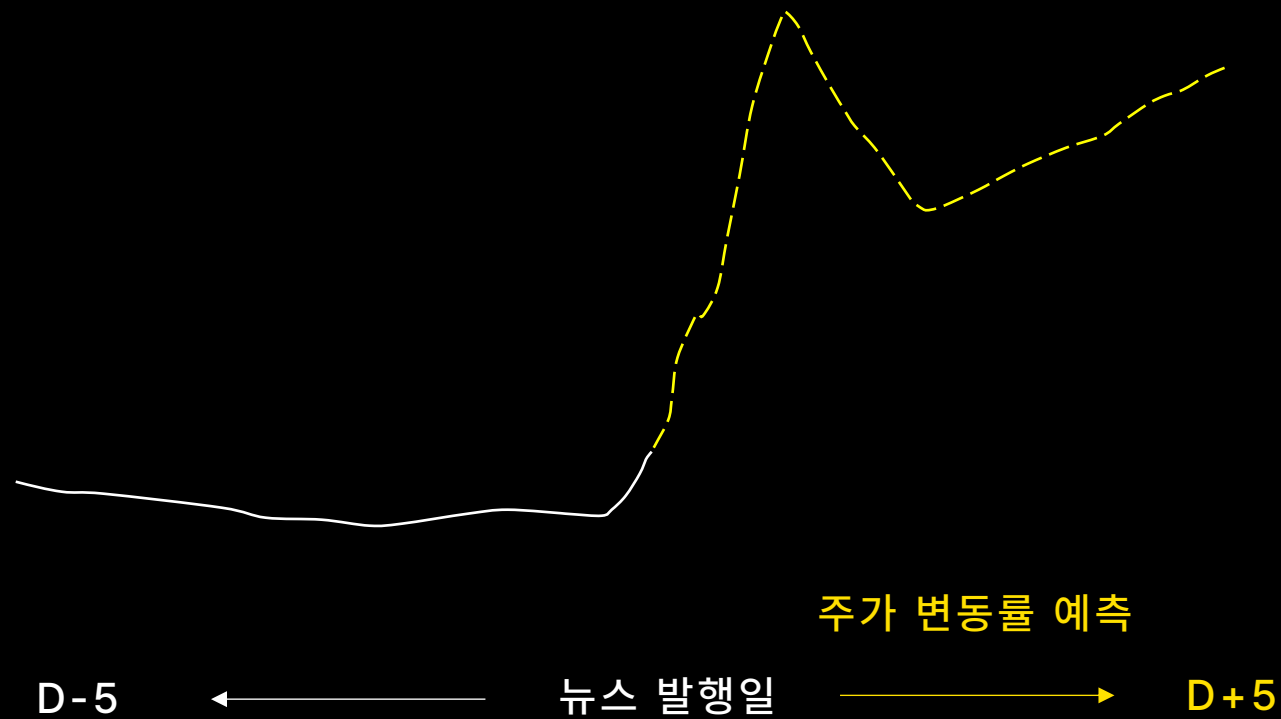


삼성바이오로직스, 바이오
헬스케어 기업의 주가
상승세

국내 바이오 헬스케어
산업이 상승기에 진입하고
있는 지금이 해당 산업의
투자 적기라고 설명했다

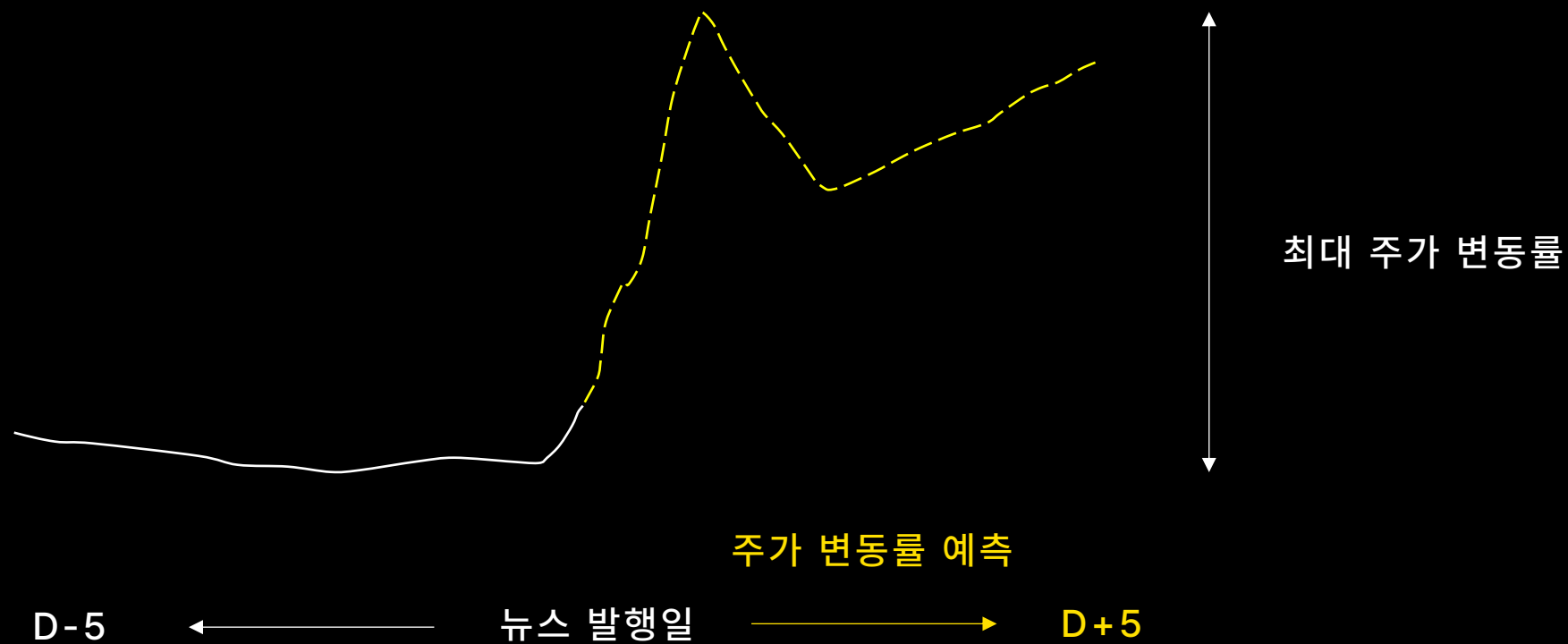


3-2 주요 뉴스 선별



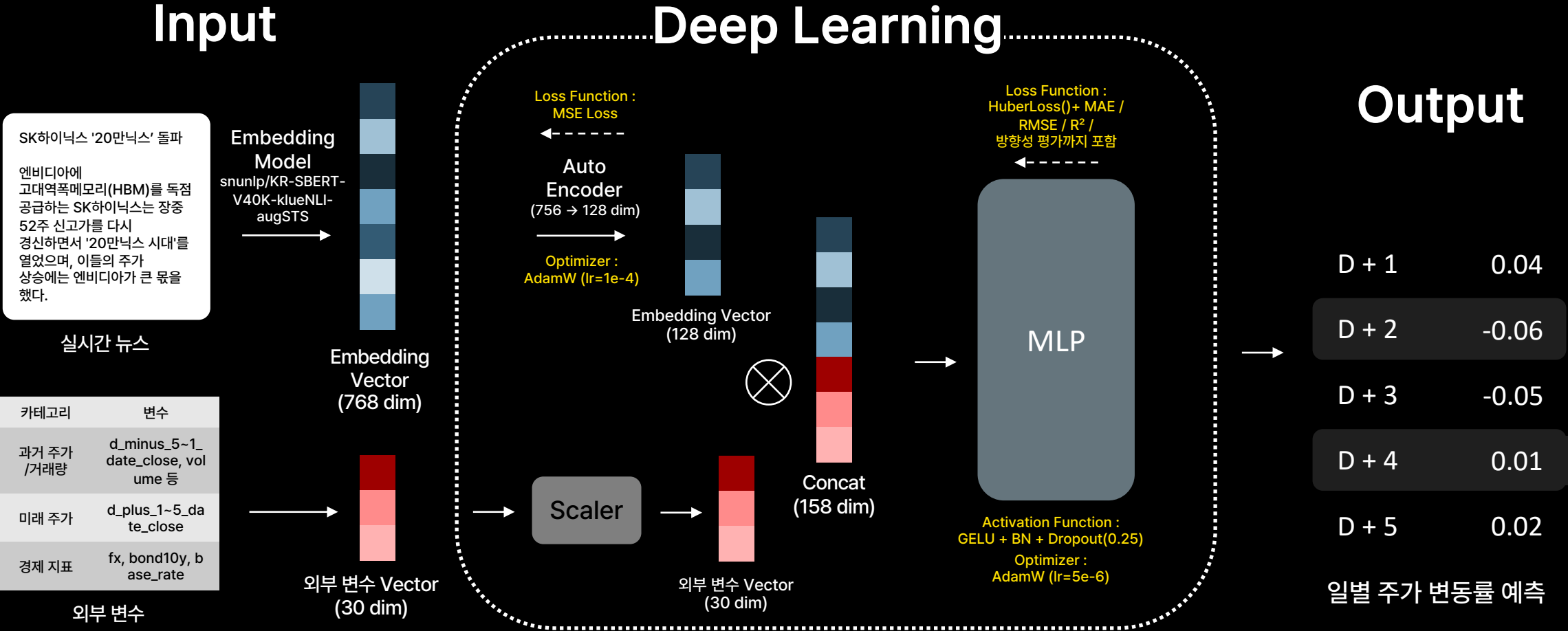
토스 증권 벤치마킹

3-2 주요 뉴스 선별



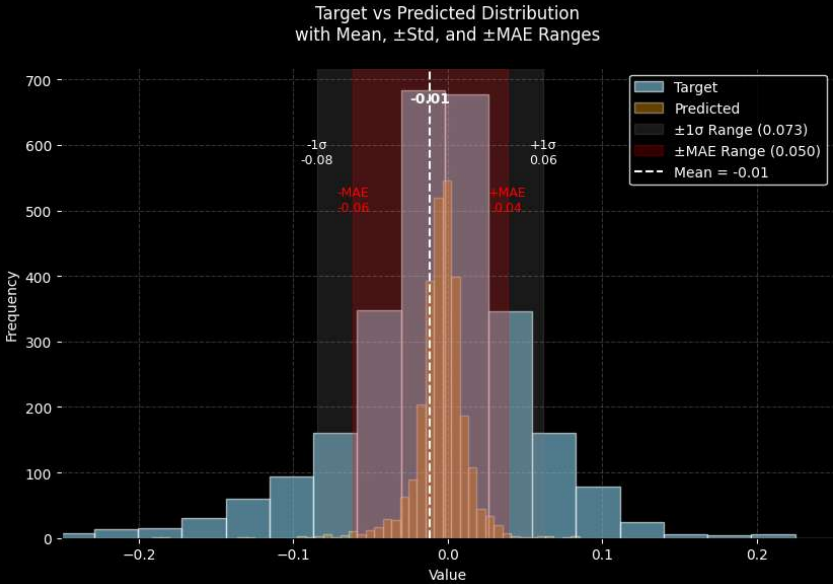
뉴스 중요도 = 뉴스 발행일 5일 전,후 최대 주가 변동률

3-2 주요 뉴스 선별



X : 뉴스 요약, 수치형 변수 y : 해당 뉴스에 매칭된 종목의 1~5일 후 주가 변동률

3-2 주요 뉴스 선별

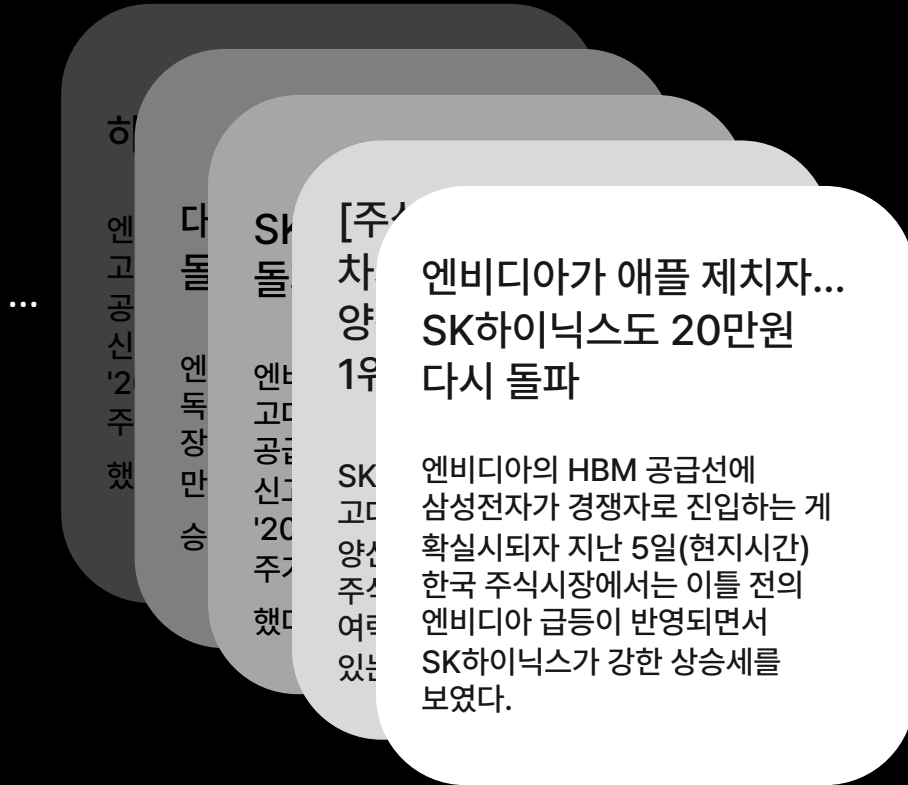


MAE: 0.05

뉴스 요약	뉴스 중요도
NH투자증권은 불닭볶음면 수출 호조로...	23
대신증권은 방산 수요 증가와 구리 강세로...	37
KB증권은 DL이앤씨가 주택 원가율 개선과...	70
KB증권은 카카오뱅크의 목표주가를 3만4000원으로...	63
NH투자증권은 롯데칠성의 2분기 실적 부진에도 불구하고...	17
신한투자증권은 화장품주에 대해 밸류에이션 랠리는 타당...	13
SK하이닉스가 내년 엔비디아 HBM4 공급의 70% 이상을...	40
최근 원·달러 환율과 유가 하향 안정화에 따른 수혜주로...	100

3-3 과거 유사 뉴스 선별

3-3 과거 유사 뉴스 선별



13,000건의 뉴스 데이터

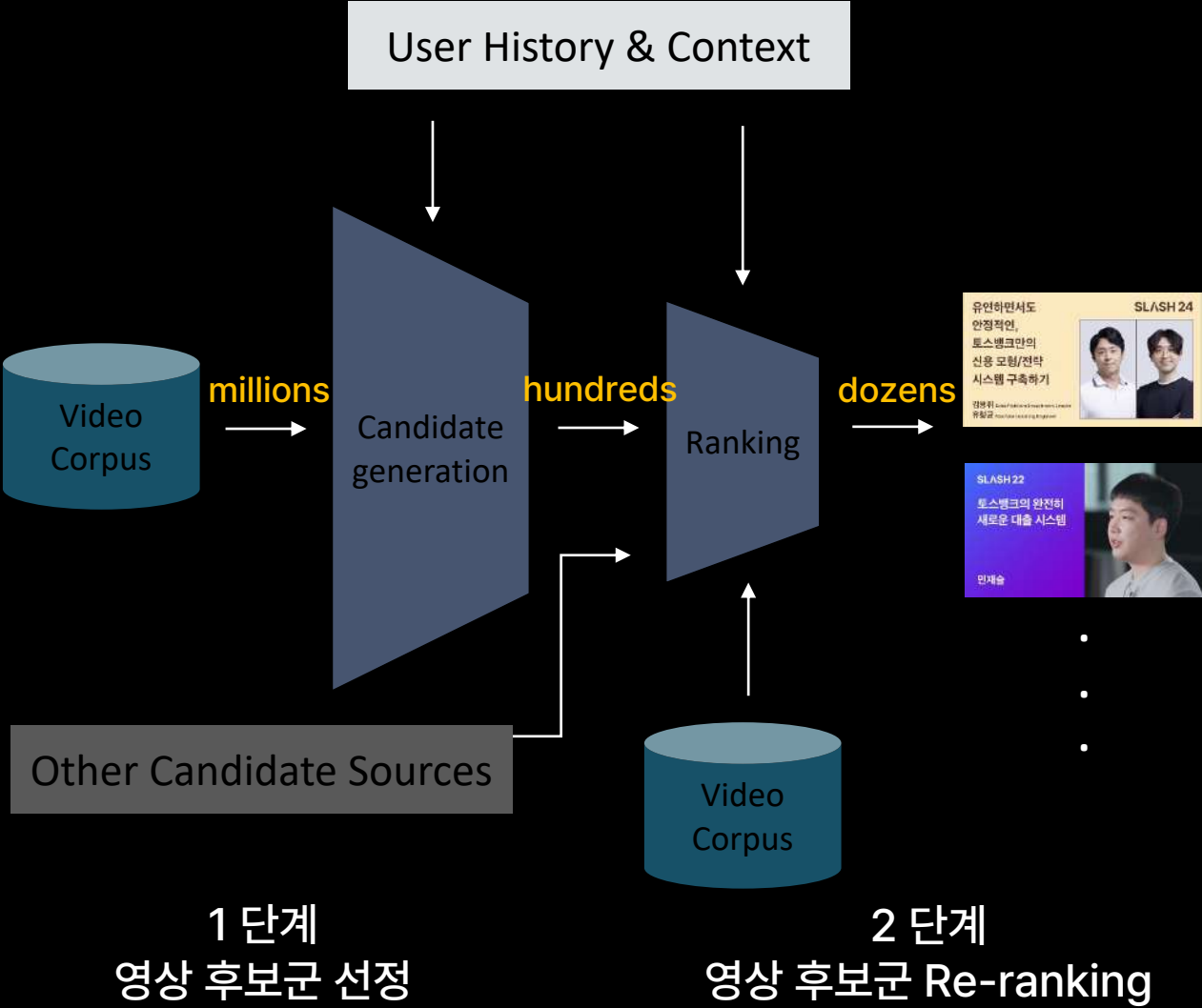
많은 뉴스 데이터에서
어떻게 유사 뉴스를 선별해야 할까?

3-3 과거 유사 뉴스 선별



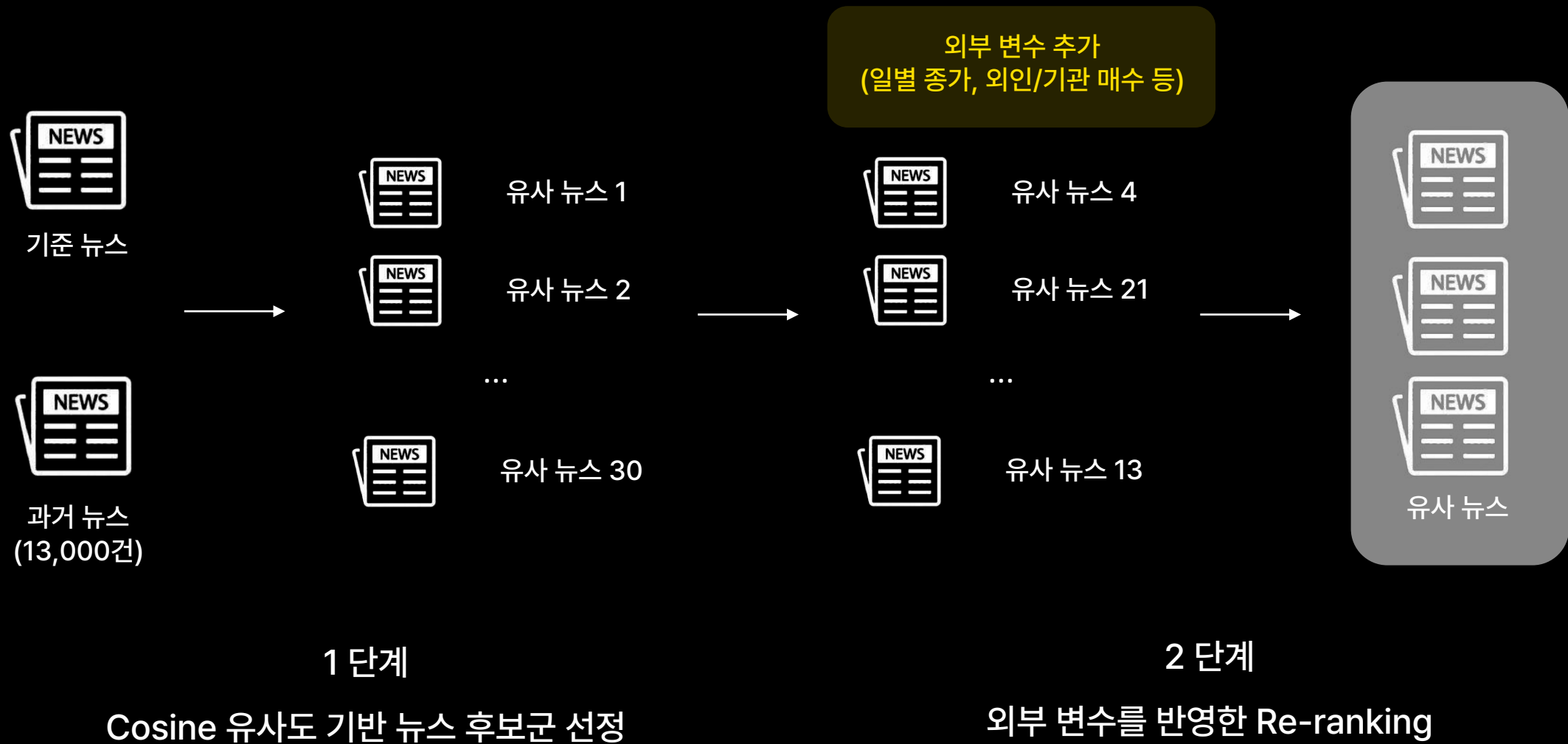
유튜브 영상 추천 시스템

Deep Neural Networks for YouTube Recommendations
(2016)

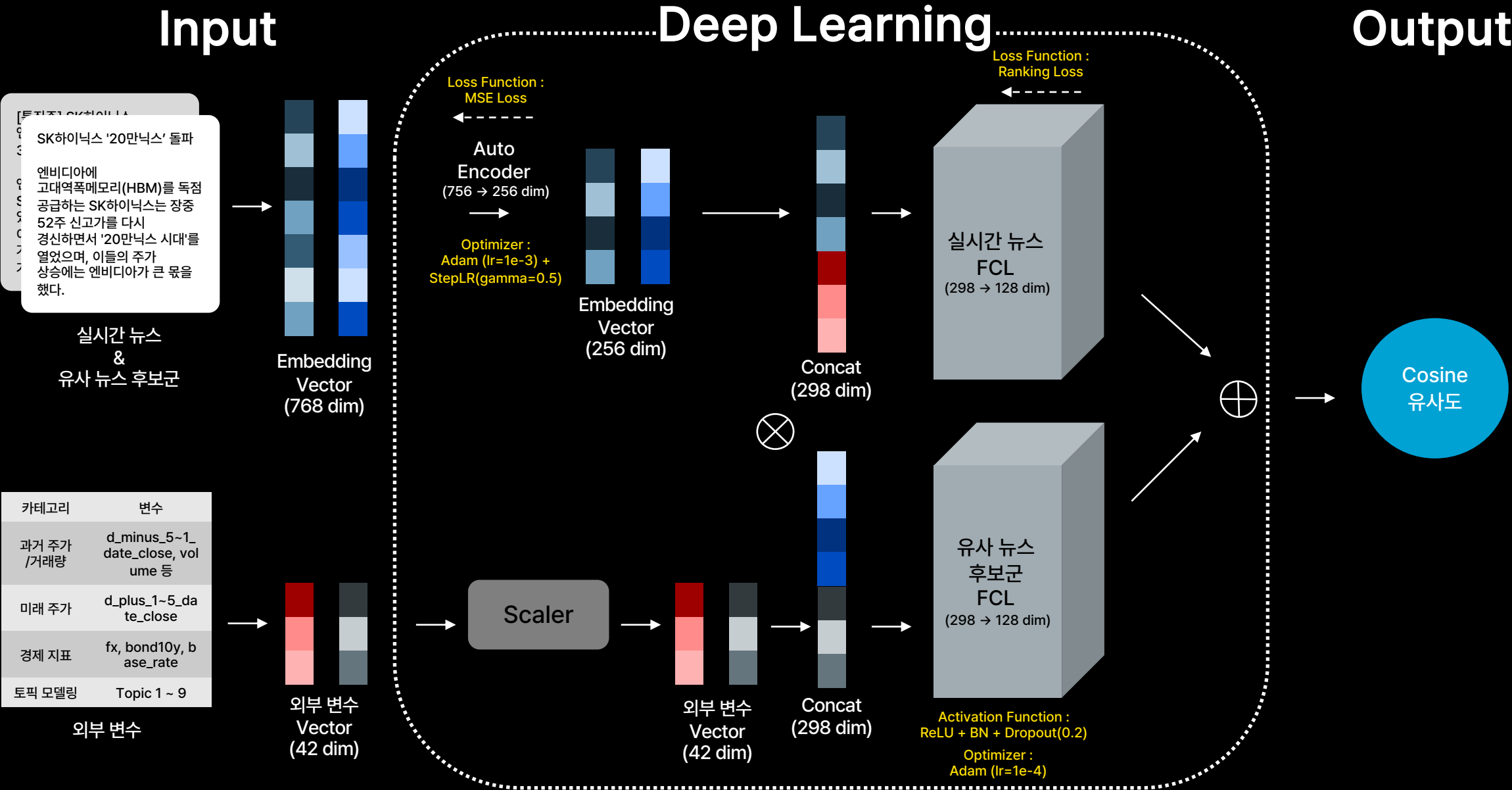


3-3 과거 유사 뉴스 선별

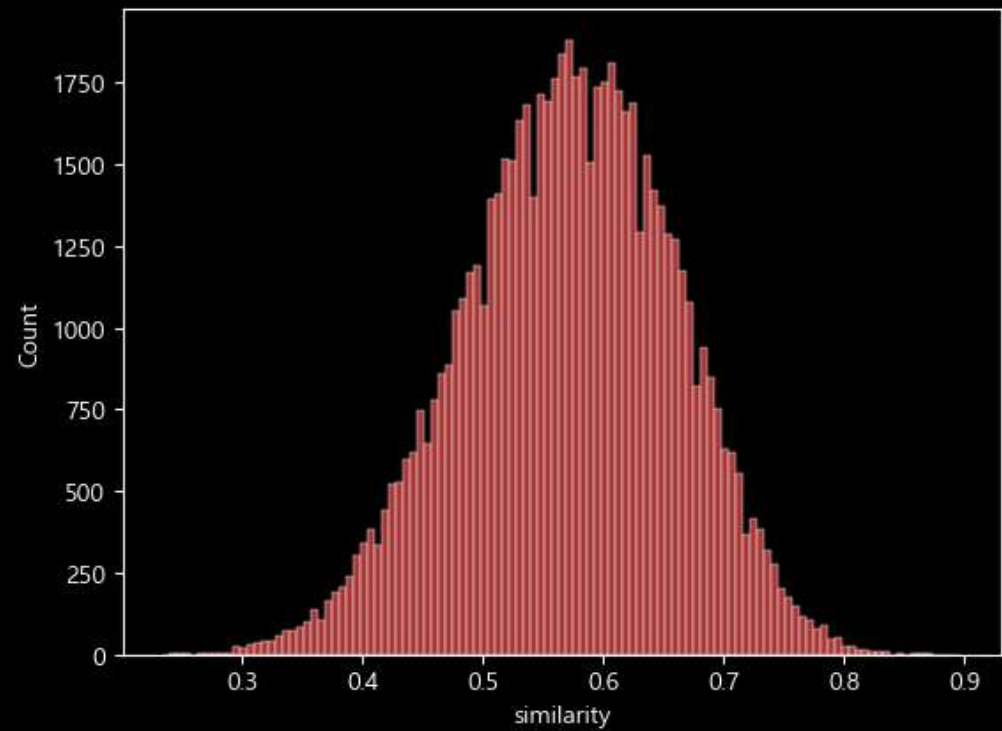
과거 유사 뉴스 선별 모델



3-3 과거 유사 뉴스 선별



3-3 과거 유사 뉴스 선별



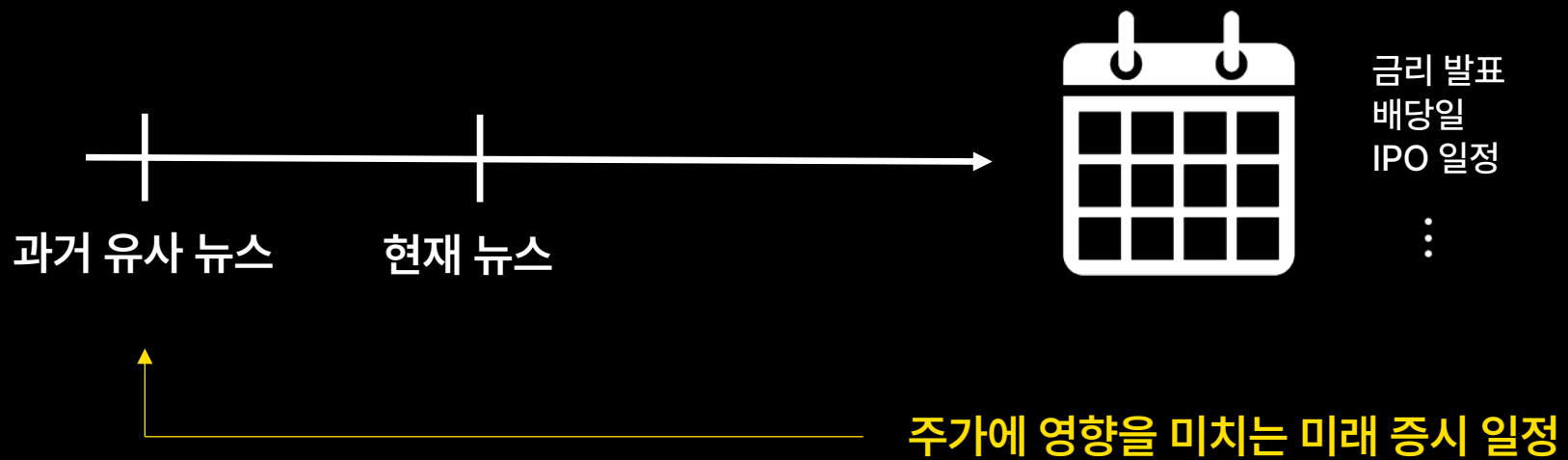
유사도 평균: 0.58

유사 뉴스	실제 유사도	예측 유사도
AI 반도체 열풍 재점화... SK하이닉스 기대감 ↑	0.76	0.68
SK하이닉스, AI 붐에 몸값 상승... 20만원 재돌파	0.70	0.52
'AI 훈풍'에 엔비디아·SK하이닉스 주가 동반 상승	0.68	0.62
엔비디아 효과... SK하이닉스·삼성전자도 웃었다	0.65	0.44
'AI 대장株' 엔비디아 질주에 한국 반도체株 들썩	0.61	0.74

Precision@5: 0.36

3-4 과거 유사 뉴스 챗봇

3-4 과거 유사 뉴스 챗봇

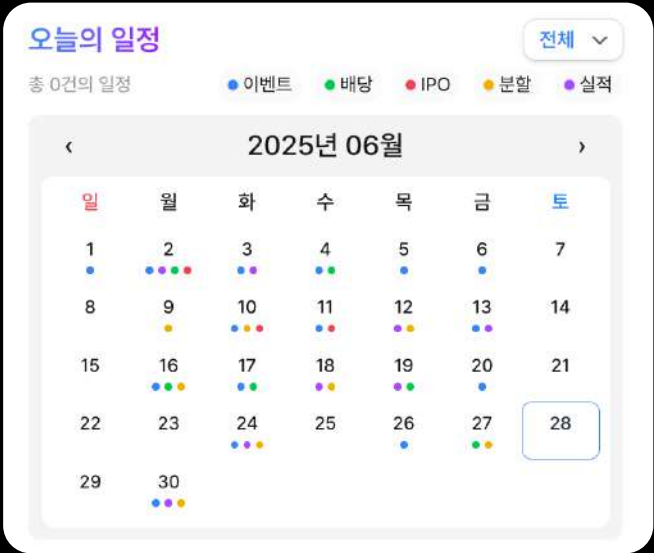


3-4 과거 유사 뉴스 챗봇

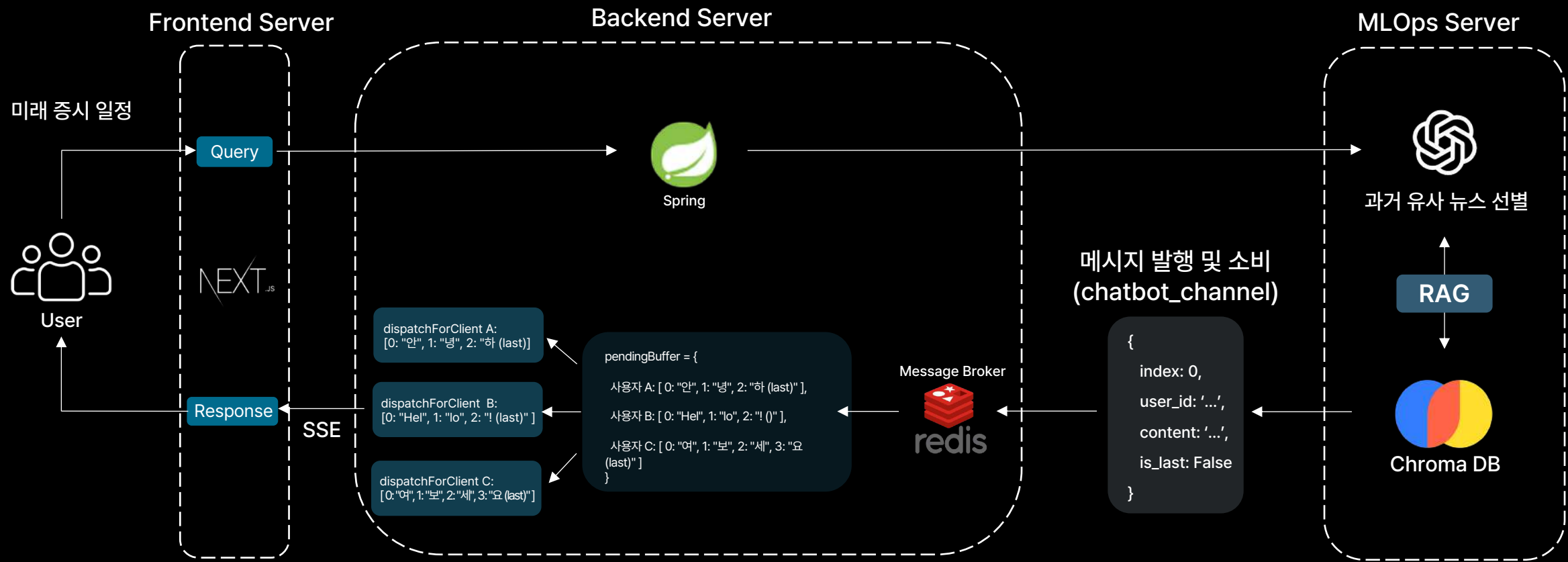


미래 증시 일정

데이터 수집



3-4 과거 유사 뉴스 챗봇



3-5 개인 맞춤 뉴스 추천

3-5 개인 맞춤 뉴스 추천

유저 데이터

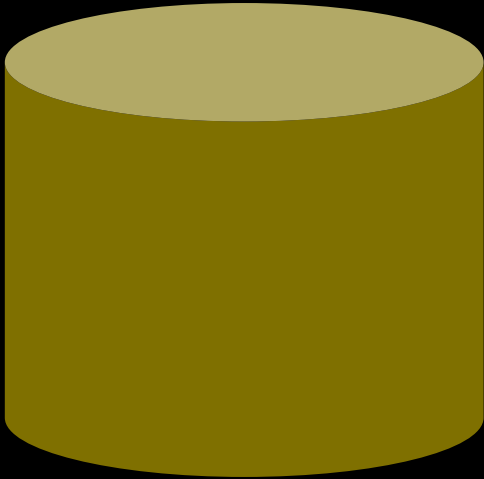
user_id	순수익	투자 성향	보유 종목
u1	1204...	1	카카오페이
u2	4323...	2	삼성전자
u3	2209...	1	한화오션
u4	5983...	3	스마일게이트, 넷마블

로그 데이터

user_id	News_id	클릭 여부
u1	n3	1
u1	n2	0
u2	n5	1
u3	n7	0

뉴스 데이터

user_id	요약	관련 종목	시장 변동률	중요도
n1	카카오 페이가 ...	1	...	45
n2	반도체 호황이 ...	2	...	85
n2	친환경 에너지 ...	1	...	32
n4	게임 업계에...	3	...	93

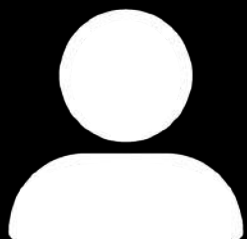


뉴스 추천 모델

수집된 데이터 기반으로 개인 맞춤 뉴스를 추천



3-5 개인 맞춤 뉴스 추천



보유 자산: 12,000,000원

순수익: 480,000원

✓ 투자 성향: 4 (공격형)

보유 종목: SK하이닉스



✓ 뉴스 중요도: 87점

카카오페이는 새 정부 출범 이후
원화 스테이블코인 도입 기대감
이 높아지면서 이달 들어 주가가
147.8% 급등했다.



✓ 추천 확률: 82%

투자 성향 고려

증권사 사업 분석

투자 수익률 고려

3-5 개인 맞춤 뉴스 추천

유저 데이터				뉴스 데이터				
보유 자산	순수익	투자 성향	종목 일치	토픽1	토픽2	...	토픽9	클릭 여부
4800...	1204...	1	1	0.453	0.456	...	0.364	0
7200...	4323...	2	0	0.476	0.875	...	0.132	1
1236...	2209...	1	0	0.123	0.432	...	0.147	1
5912...	5983...	3	1	0.687	0.569	...	0.298	1

200,000 row

학습 데이터 구성

- 토픽1 : 글로벌 시장 동향
토픽2 : 증권사 주가 전망
토픽3 : 기업 실적 개선
토픽4 : 증권사 사업 분석
토픽5 : 대기업 사업 전략
토픽6 : 기업 기술 개발
토픽7 : 반도체 및 AI
토픽8 : 금융 서비스
토픽9 : 주주 및 경영 이슈

뉴스 토픽별 키워드

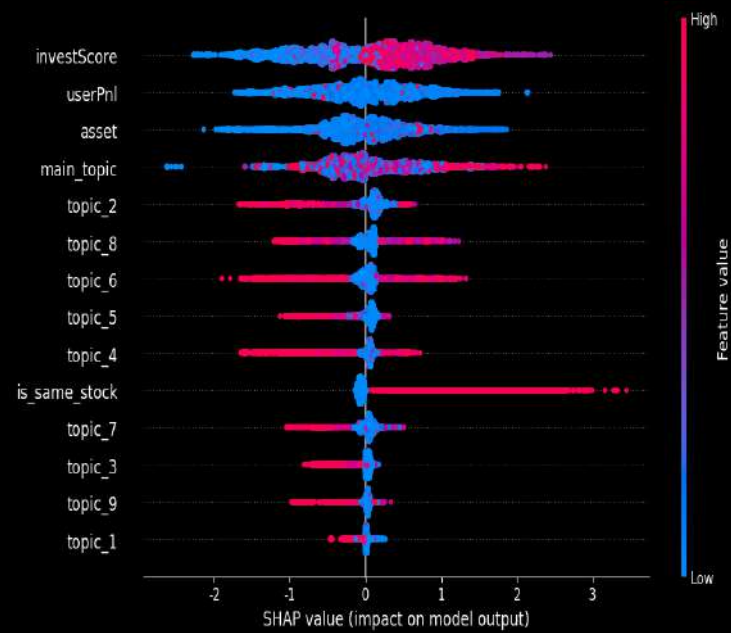
뉴스 클릭 여부 기반 ML 분류 모델링



3-5 개인 맞춤 뉴스 추천

LightGBM

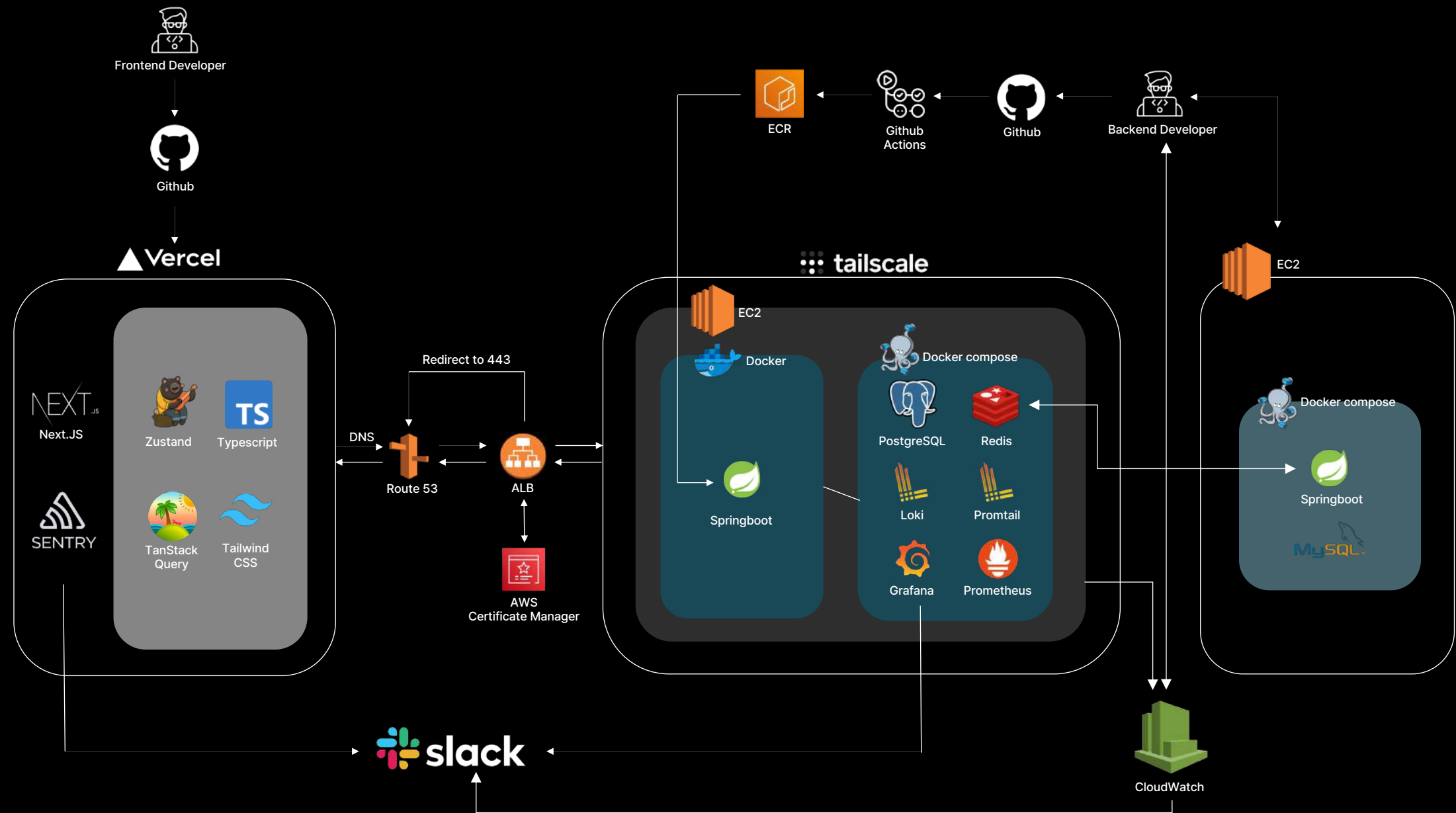
Precision@5: 0.46



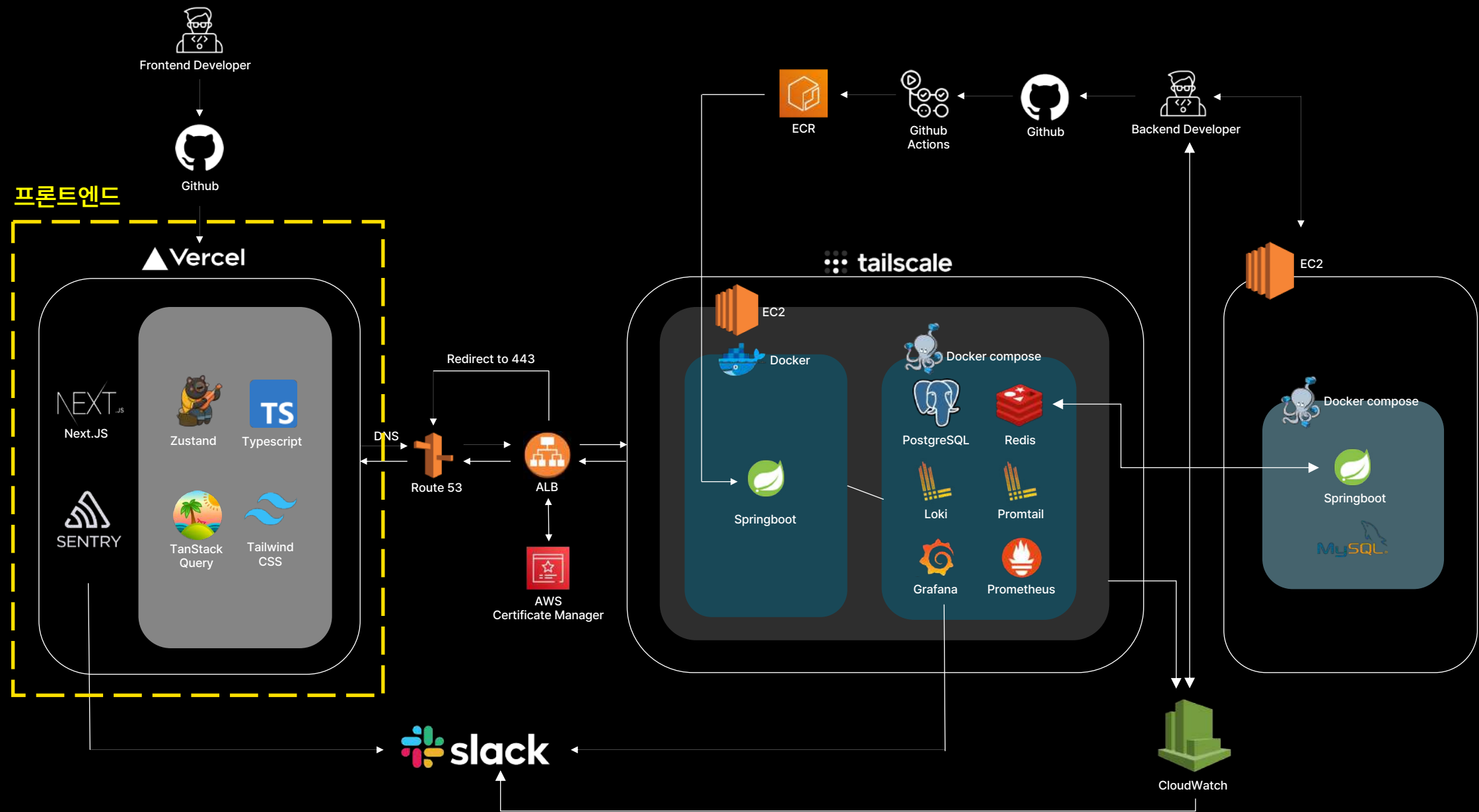
추천 뉴스	추천 확률	추천 이유	...
AI 반도체 열풍 재점화... SK하이닉스 기대감 ↑	0.76	글로벌...	...
SK하이닉스, AI 붐에 몸값 상승... 20만원 재돌파	0.70	기업 실적...	...
'AI 훈풍'에 엔비디아·SK하이닉스 주가 동반 상승	0.68	증권사...	...
엔비디아 효과... SK하이닉스·삼성전자도 웃었다	0.65	반도체...	...
'AI 대장株' 엔비디아 질주에 한국 반도체株 들썩	0.61	대기업...	...

4 아키텍처

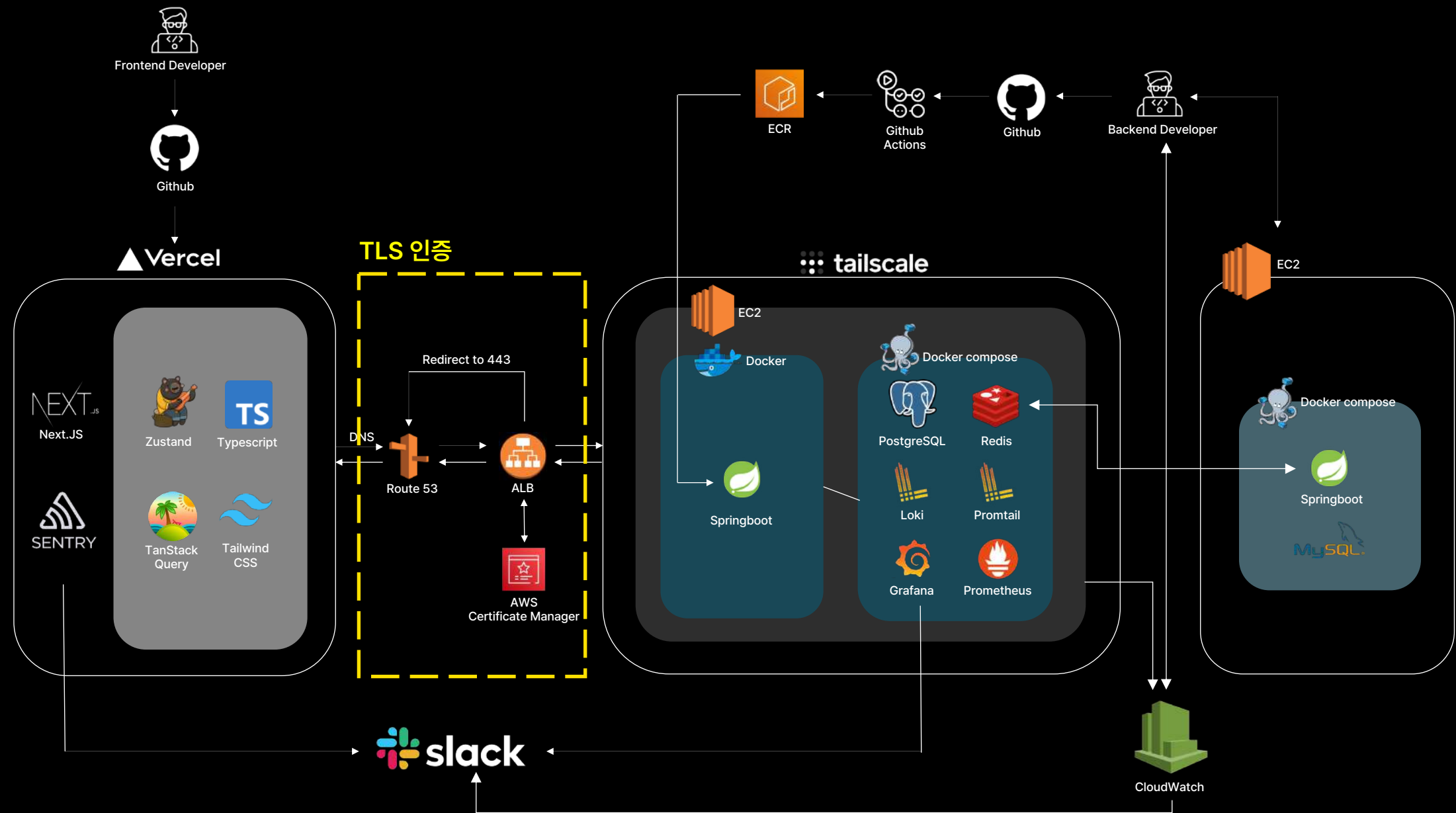
4-1 FullStack 아키텍처



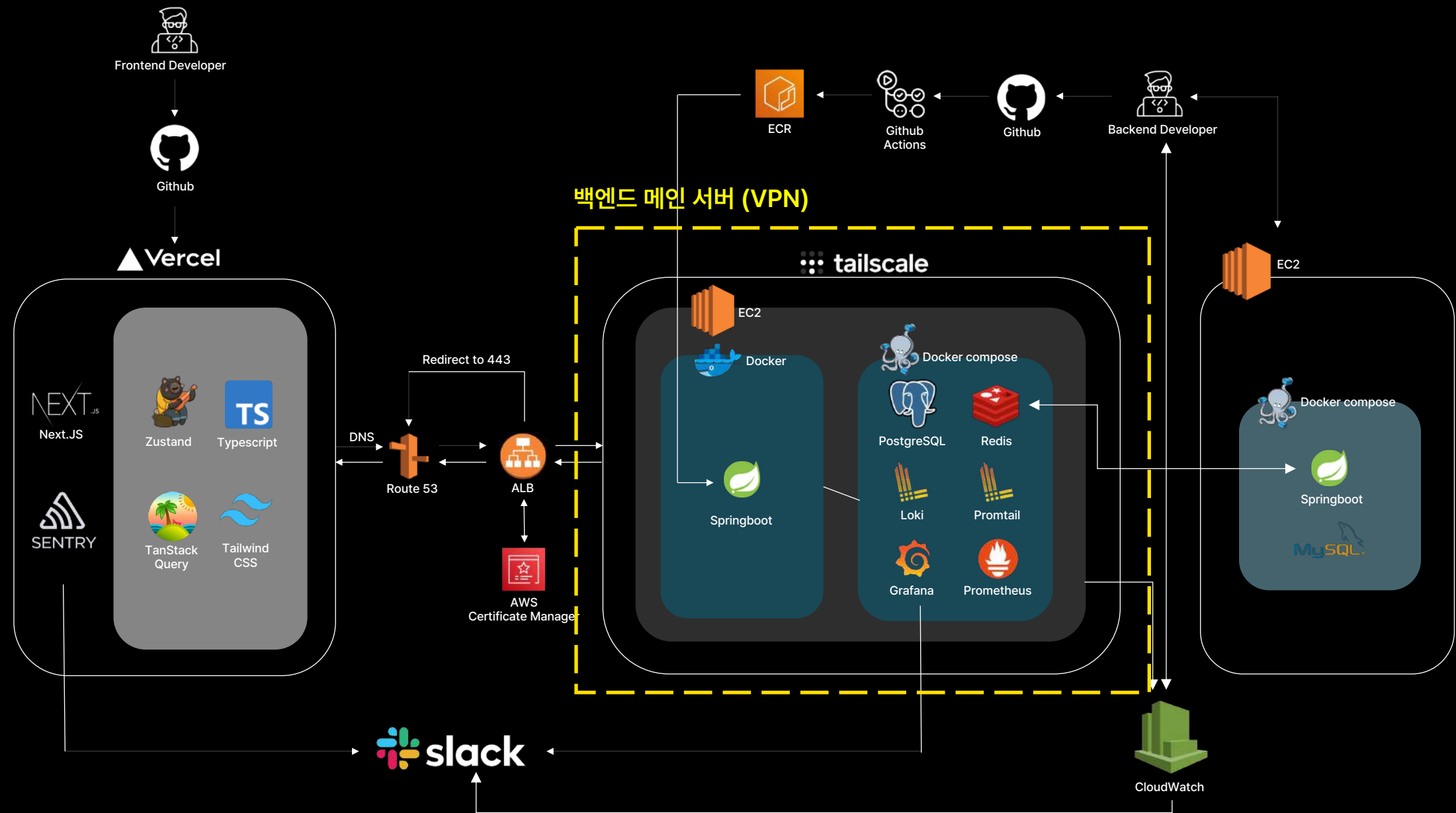
4-1 FullStack 아키텍처



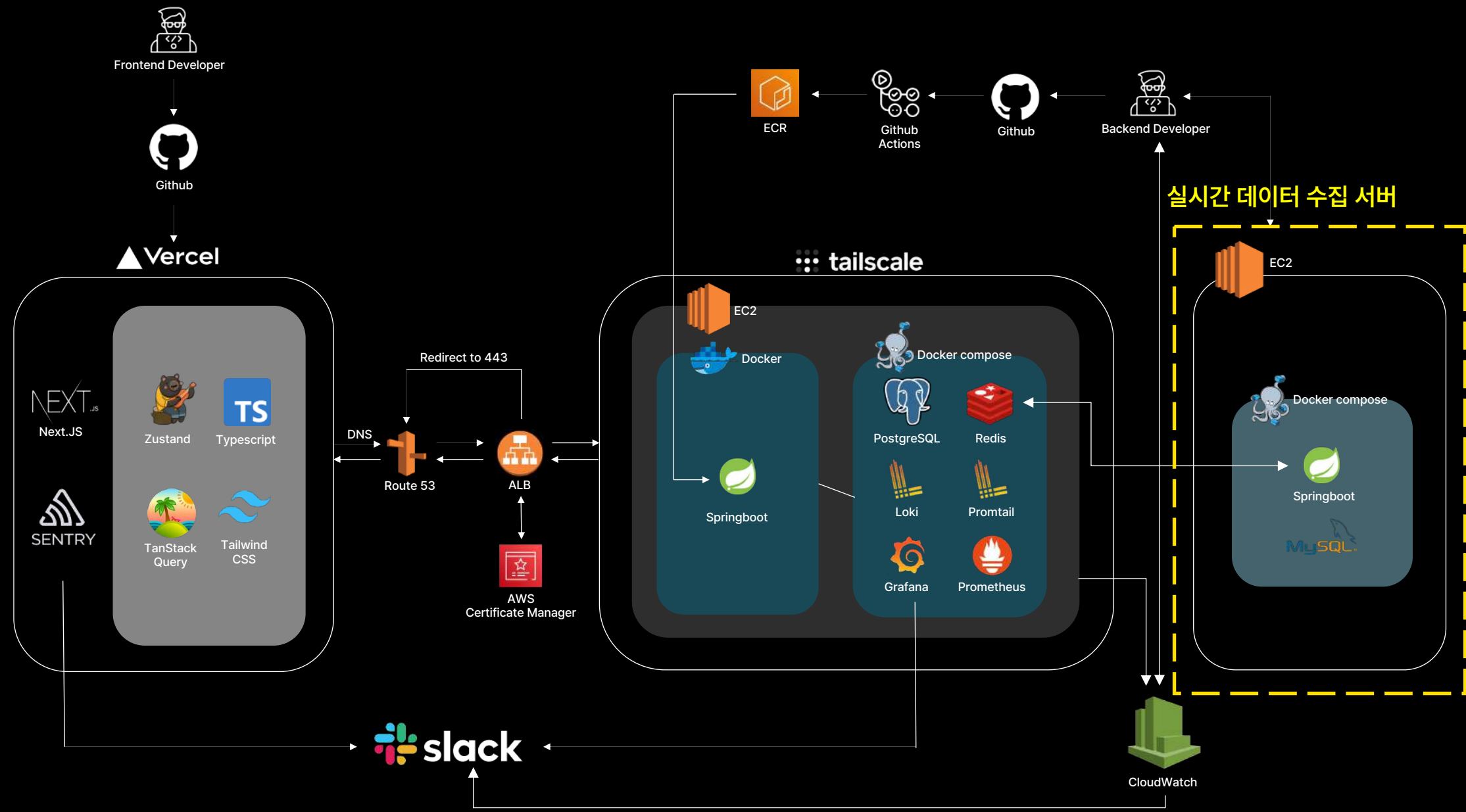
4-1 FullStack 아키텍처



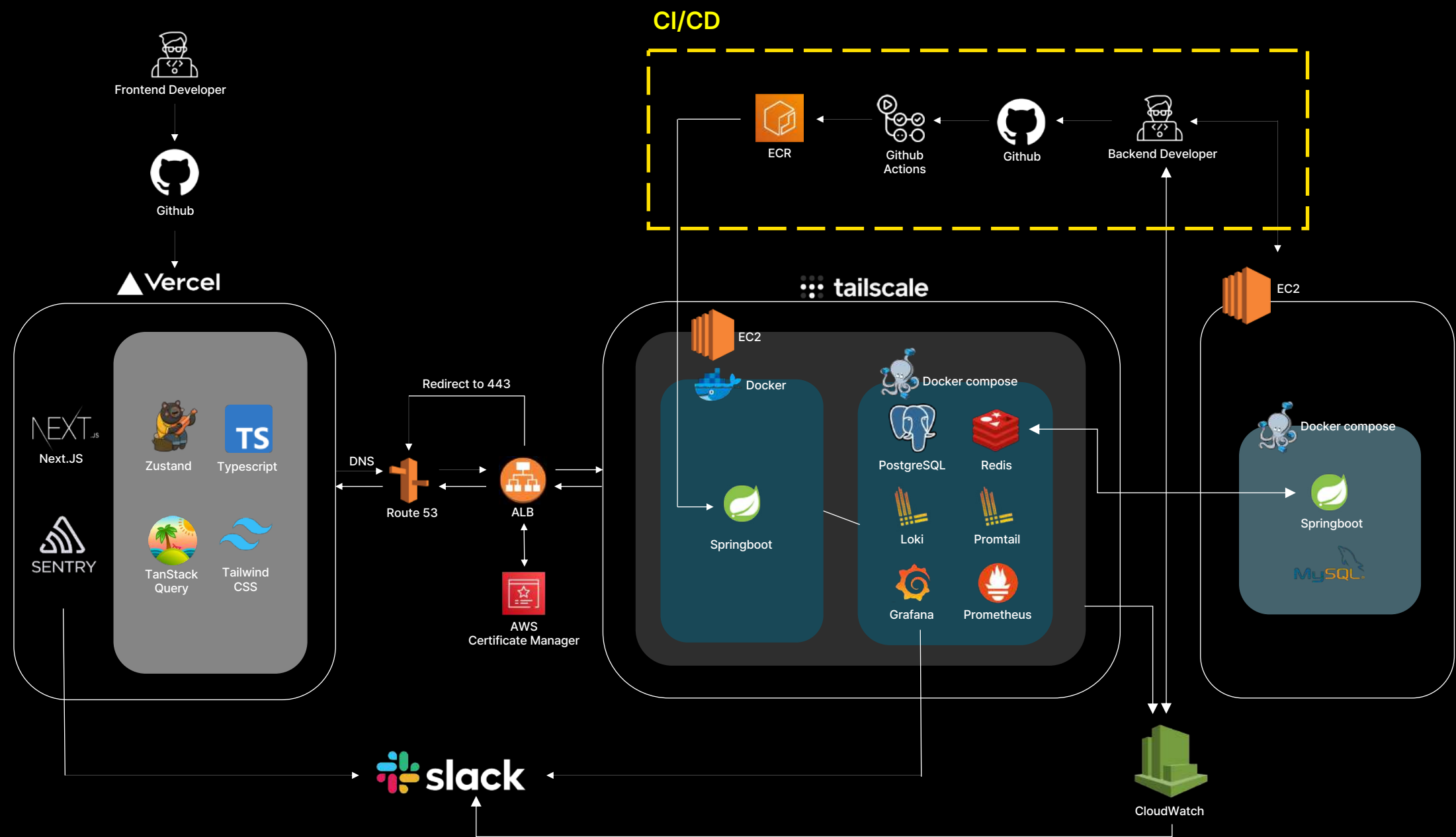
4-1 FullStack 아키텍처



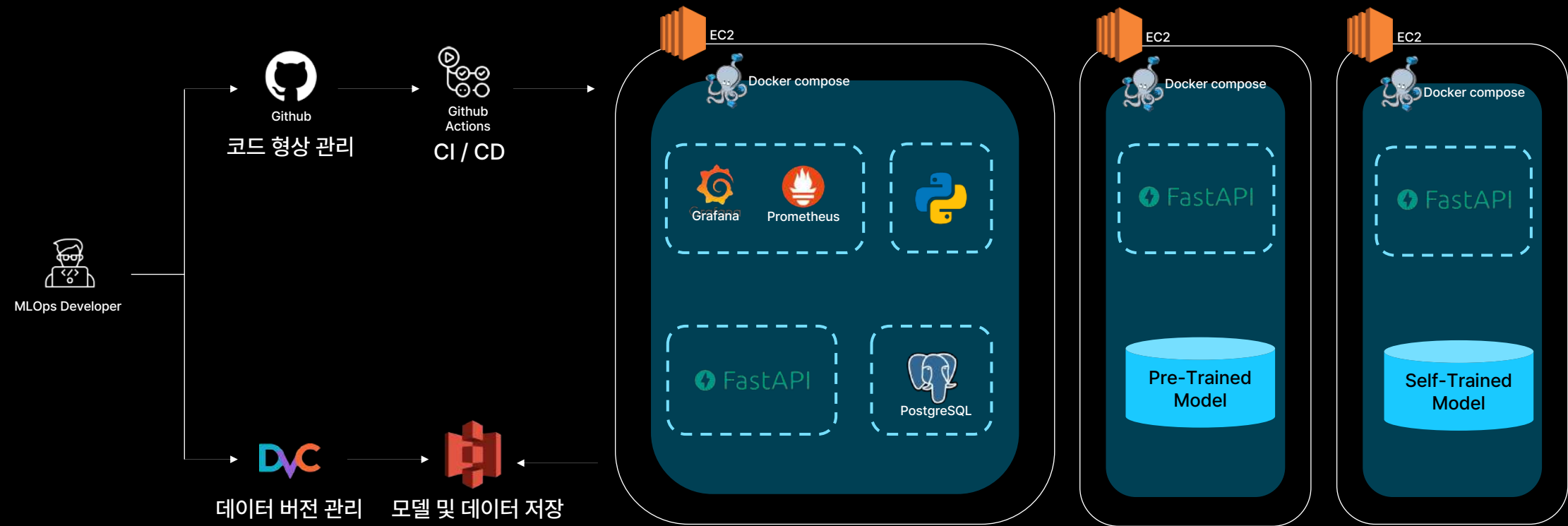
4-1 FullStack 아키텍처



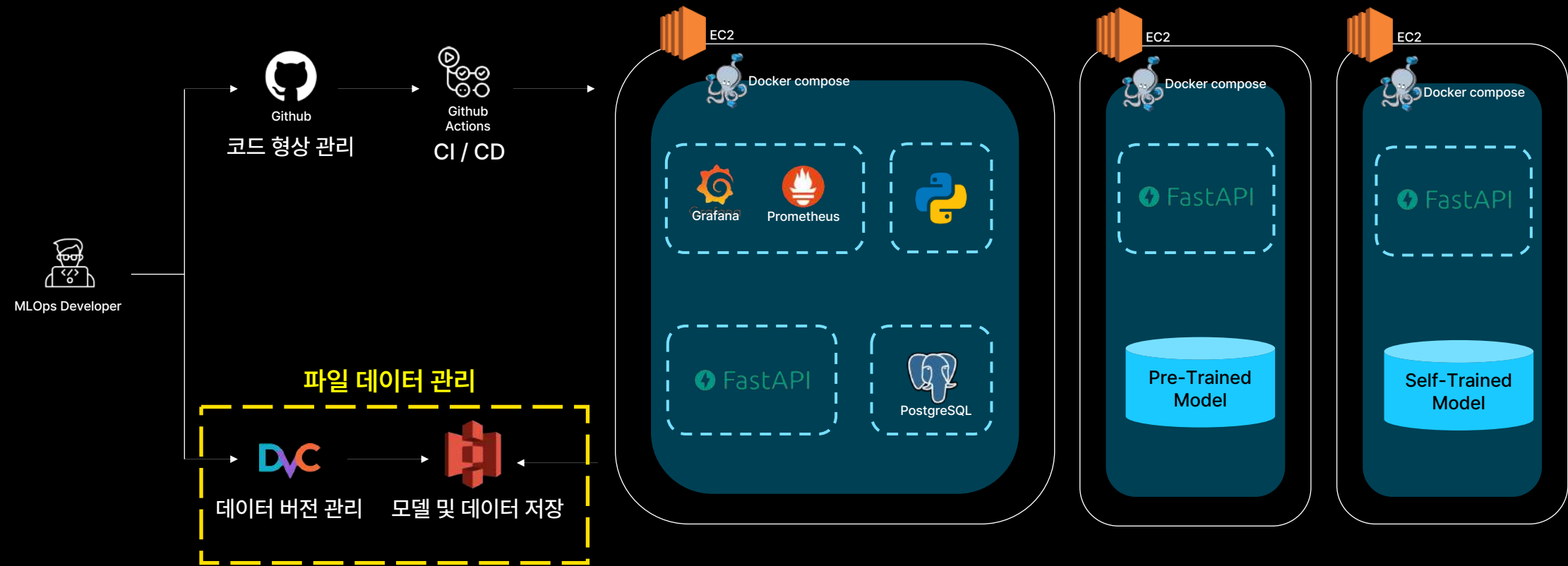
4-1 FullStack 아키텍처



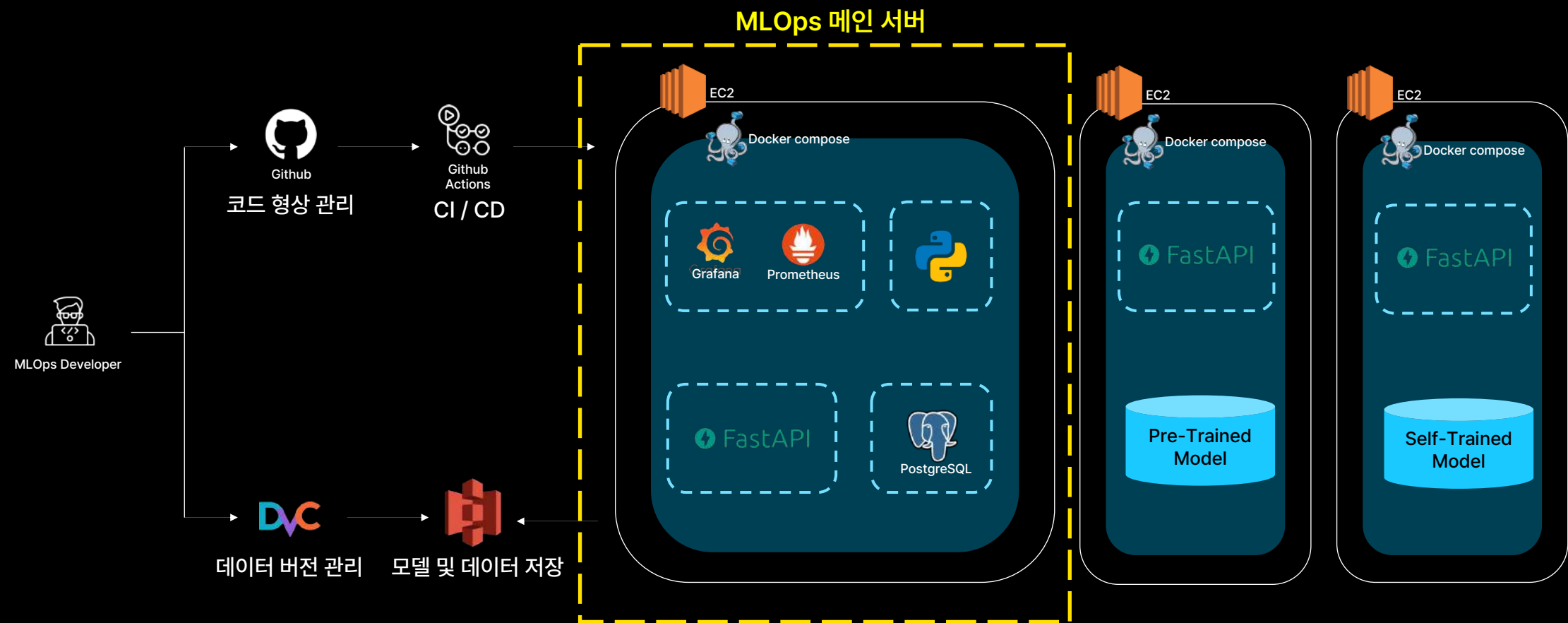
4-2 MLOps 아키텍처



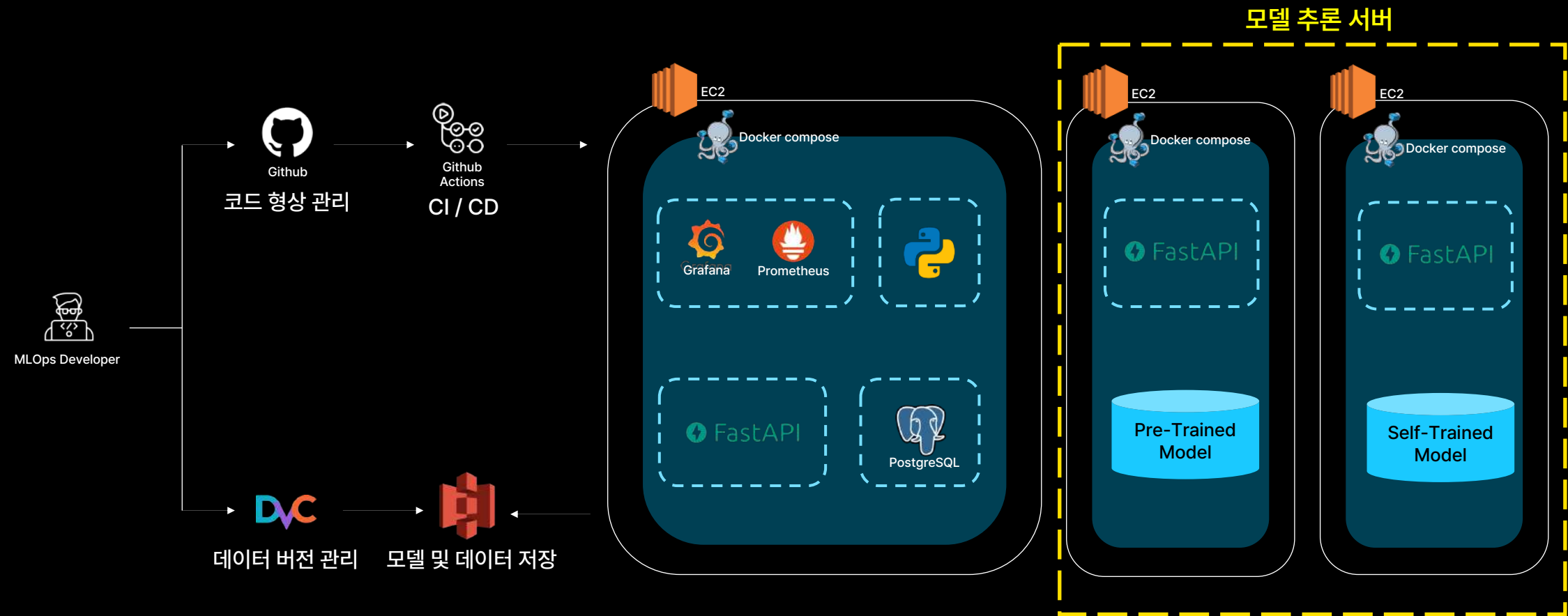
4-2 MLOps 아키텍처



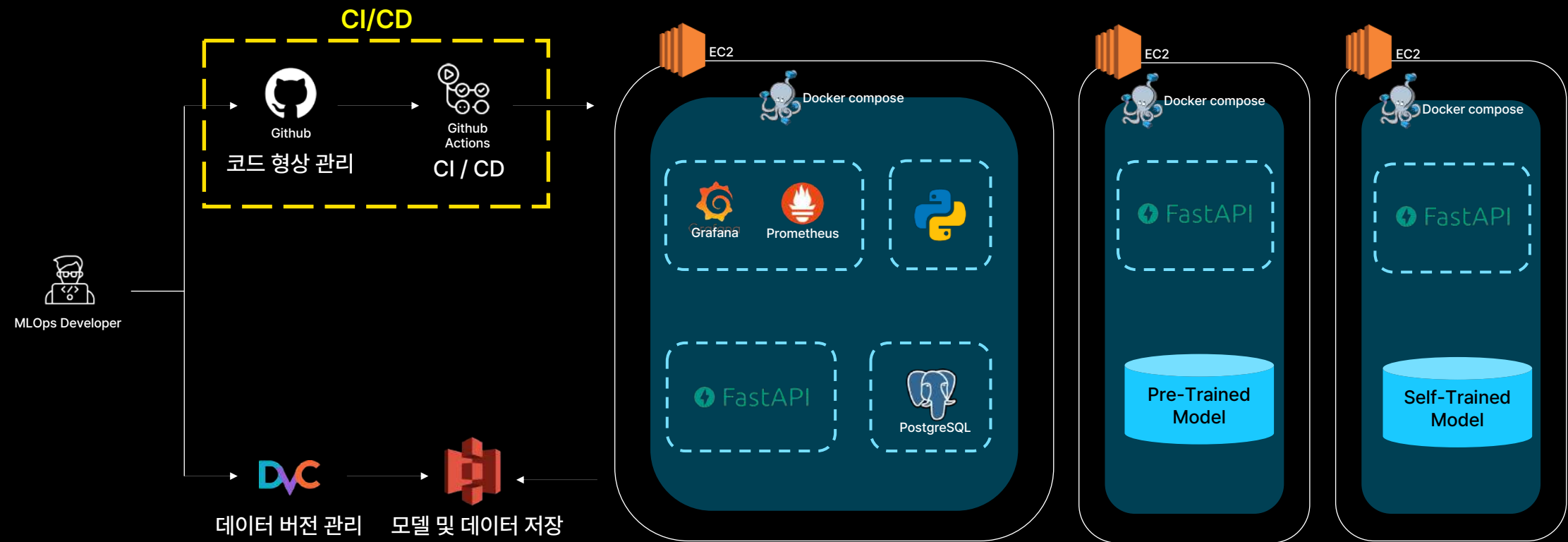
4-2 MLOps 아키텍처



4-2 MLOps 아키텍처



4-2 MLOps 아키텍처



5 트러블 슈팅

5-1 프론트엔드

5-1 Frontend 트러블 슈팅

SSR 방식 (fetch + no-store)

+

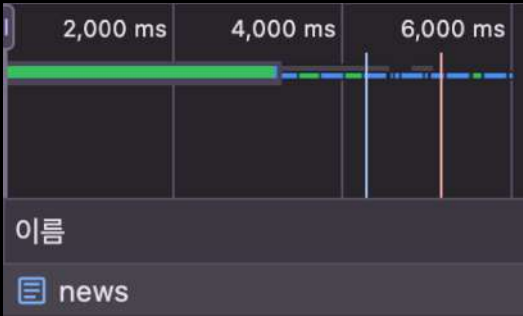
API 동기 직렬 처리

→

페이지 렌더링 속도
약 3000ms



```
1 const res1 = await fetch();
2 const res2 = await fetch();
3 const res3 = await fetch();
```



뉴스 상세 페이지 로딩 지연 - SSR과 API 순차 호출로 렌더링 시간 증가

5-1 Frontend 트러블 슈팅

ISR 방식 (fetch + **revalidate**)

+

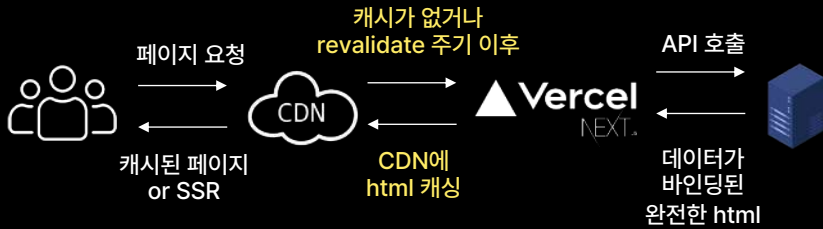
API 비동기 **병렬** 처리

+

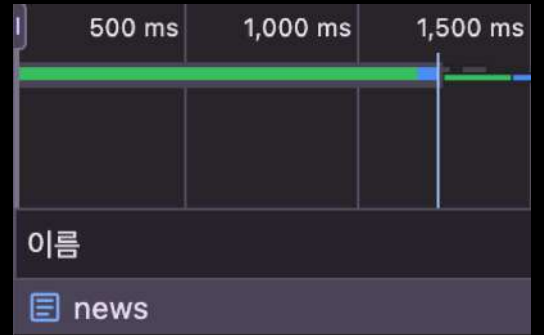
Tanstack- Query
(staleTime, gcTime)

→

페이지 렌더링 속도
약 1250ms



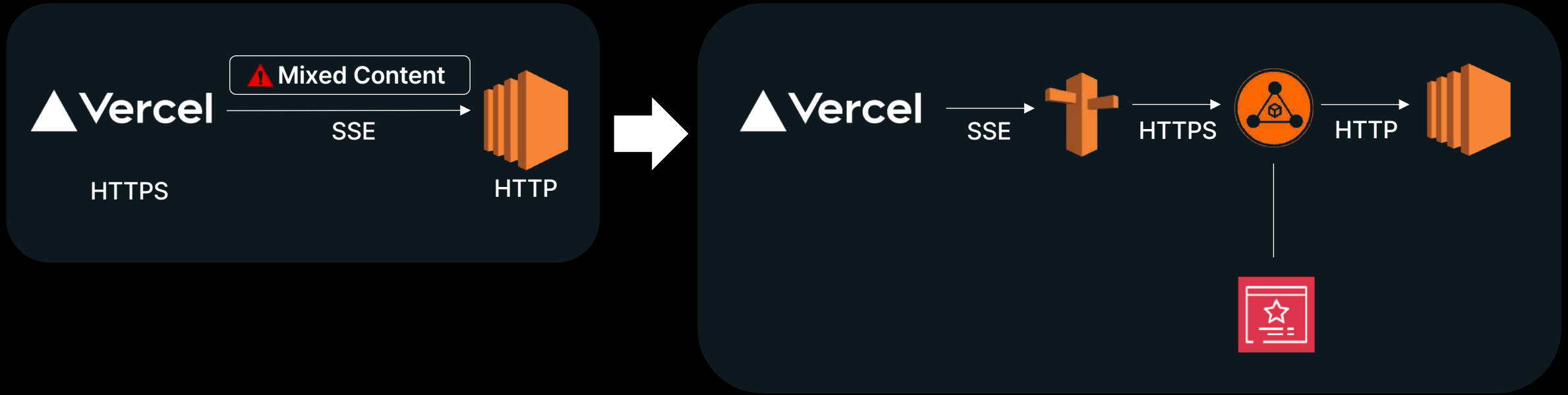
```
1 const [res1, res2, res3] = await Promise.allSettled([  
2   fetch(),  
3   fetch(),  
4   fetch(),  
5 ])
```



백그라운드 캐싱, API 병렬 처리, Tanstack-query 캐싱 적용 → **렌더링 시간 대폭 감소**

5-2 백엔드

5-2 Backend 트러블 슈팅 (1)



SSE 연결 중 Mixed Content 문제 발생 → **ALB 배치 및 SSL, TLS 인증서 적용**
외부적으로는 EC2 백엔드도 HTTPS로 통신하도록 구성

5-2 Backend 트러블 슈팅 (2)

증권 데이터 실시간 수집

CloudWatch app 오후 1:19
[🔥 CloudWatch 알람 발생 🔥]

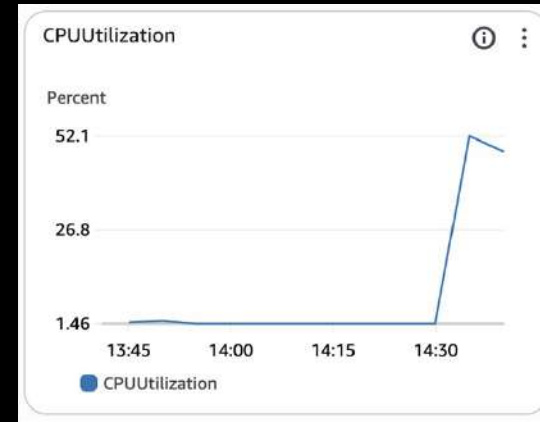
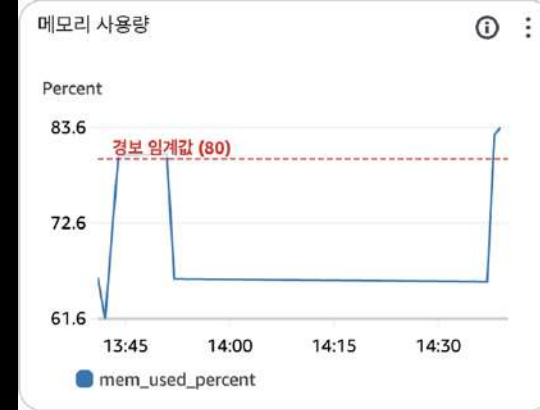
🔔 **알람 이름:** newstoss-stock-Memory-alert

📊 **상태 변경:** ALARM

📄 **설명:** # newstoss stock service의 메모리 경보입니다.
메모리가 부족하니 즉시 확인하세요

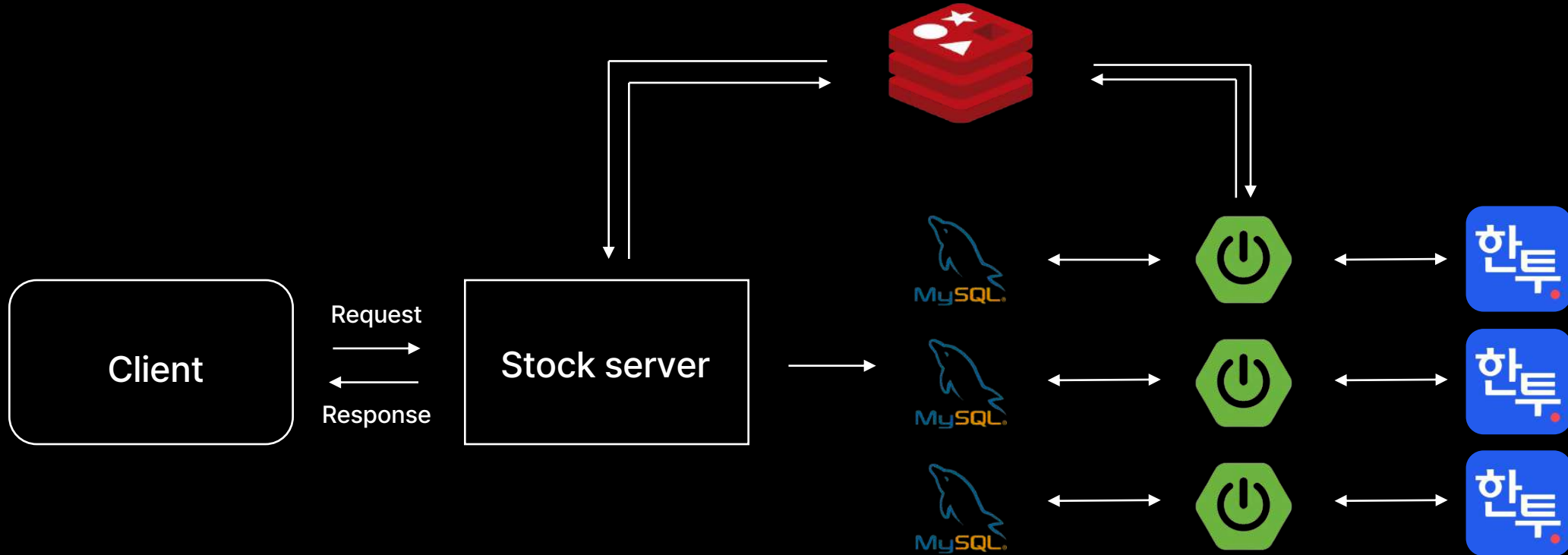
📢 **원인:**

Threshold Crossed: 1 out of the last 1 datapoints [62.69508015277197 (29/06/25 04:18:00)] was greater (minimum 1 datapoint for OK -> ALARM transition).



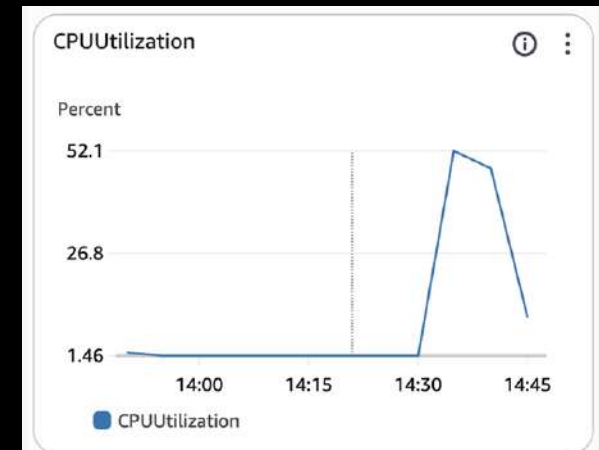
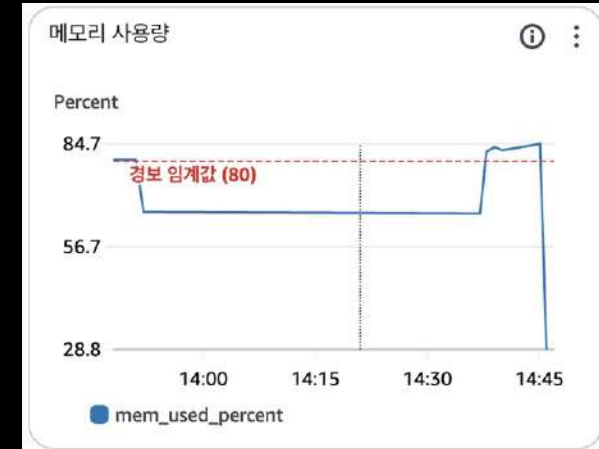
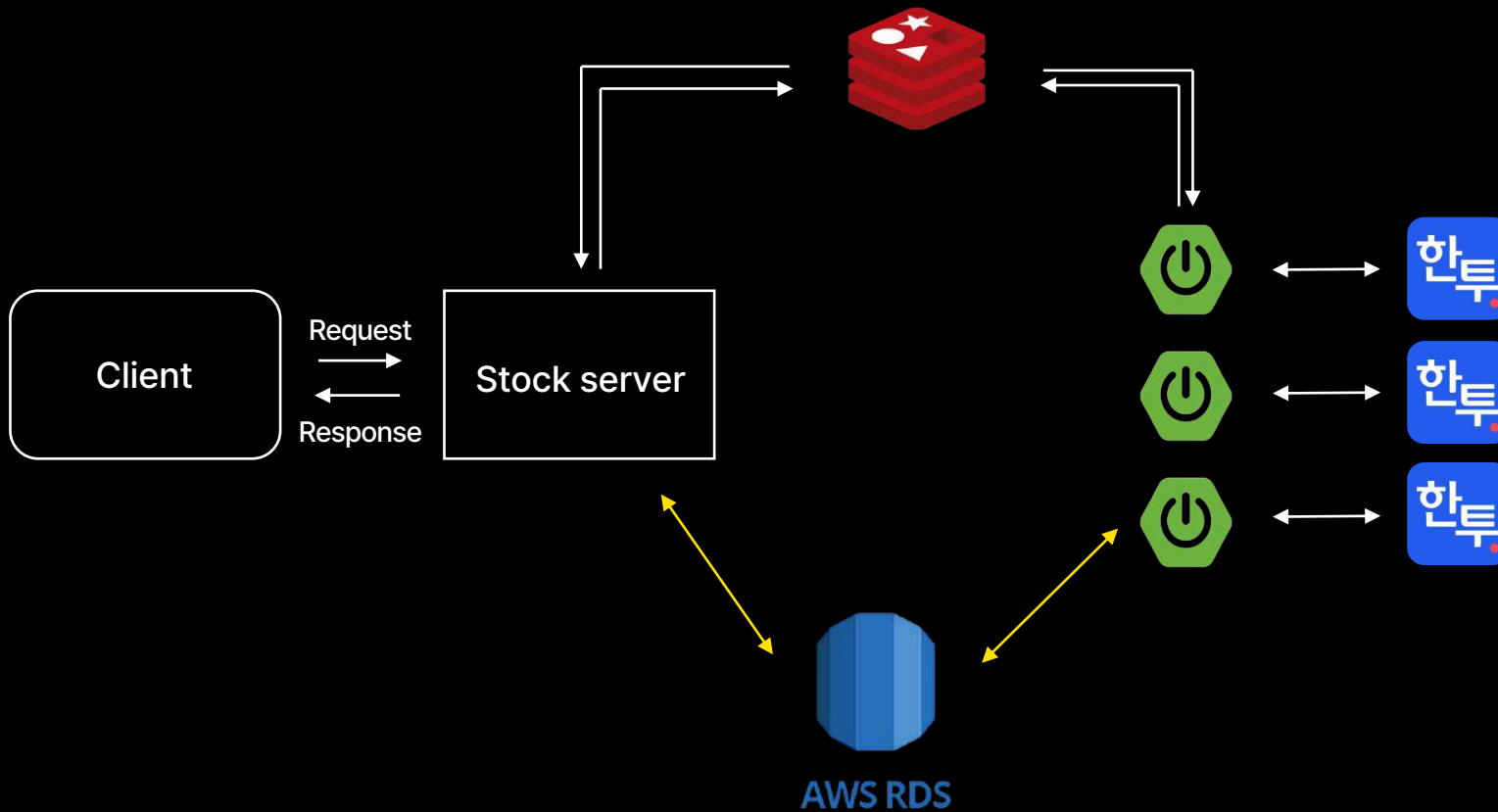
알림을 통한 서버 이슈 인지

5-2 Backend 트러블 슈팅 (2)



초기에는 DB와 Spring 서버를 동일 EC2 내 컨테이너에서 운영

5-2 Backend 트러블 슈팅 (2)

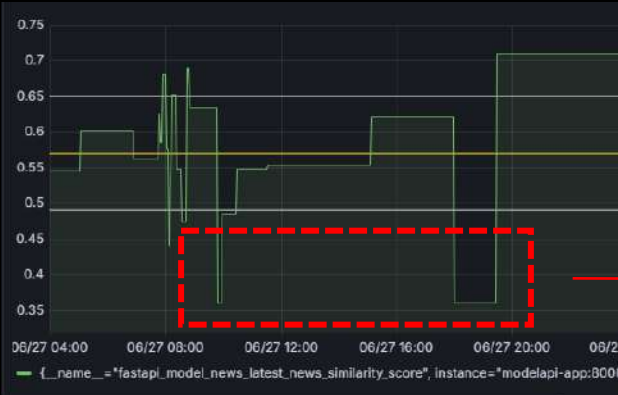


EC2에 있던 DB를 RDS로 이동 → CPU & memory 공간 확보

6 디벨롭

모니터링을 통한 모델 학습 자동화

과거 유사 뉴스 top-5 평균



평균: 0.57
편차: 0.08

이상치 탐지
평균: 0.36

낮은 유사도 평균

기준 뉴스	유사 뉴스	유사도
고려아연 경영권 분쟁 들썩...	MBK·영풍, 고려아연 공개...	0.38
고려아연 경영권 분쟁 들썩...	고려아연, 상한가 '불기둥'...	0.36
고려아연 경영권 분쟁 들썩...	한진칼, '점상'...추격매수냐...	0.35
고려아연 경영권 분쟁 들썩...	법원 가처분 기각에 고려...	0.33
고려아연 경영권 분쟁 들썩...	"영풍그룹 아닌데 급등"...	0.32

챗봇 메시지 연속성 보장 – 맞춤 뉴스 추천 고도화 및 UX 향상

Message Broker



Pub/Sub



사용자 증가



Redis Stream

⋮

MLOps

최광일

MLOps
실시간 데이터 처리

한예은

과거 유사 뉴스 선별 모델링
뉴스 관련 주식 종목 매칭

이주안

챗봇 설계
맞춤 뉴스 추천 모델링

정혜진

주요 뉴스 선별 모델링
맞춤 뉴스 추천 모델링

FullStack

박준영

뉴스 API
유저 관리 API

전장우

DevOps
캘린더 및 관심종목

박형준

증권 서비스 및 포트폴리오
모니터링

김우곤

프론트엔드 담당
UX / UI 구성

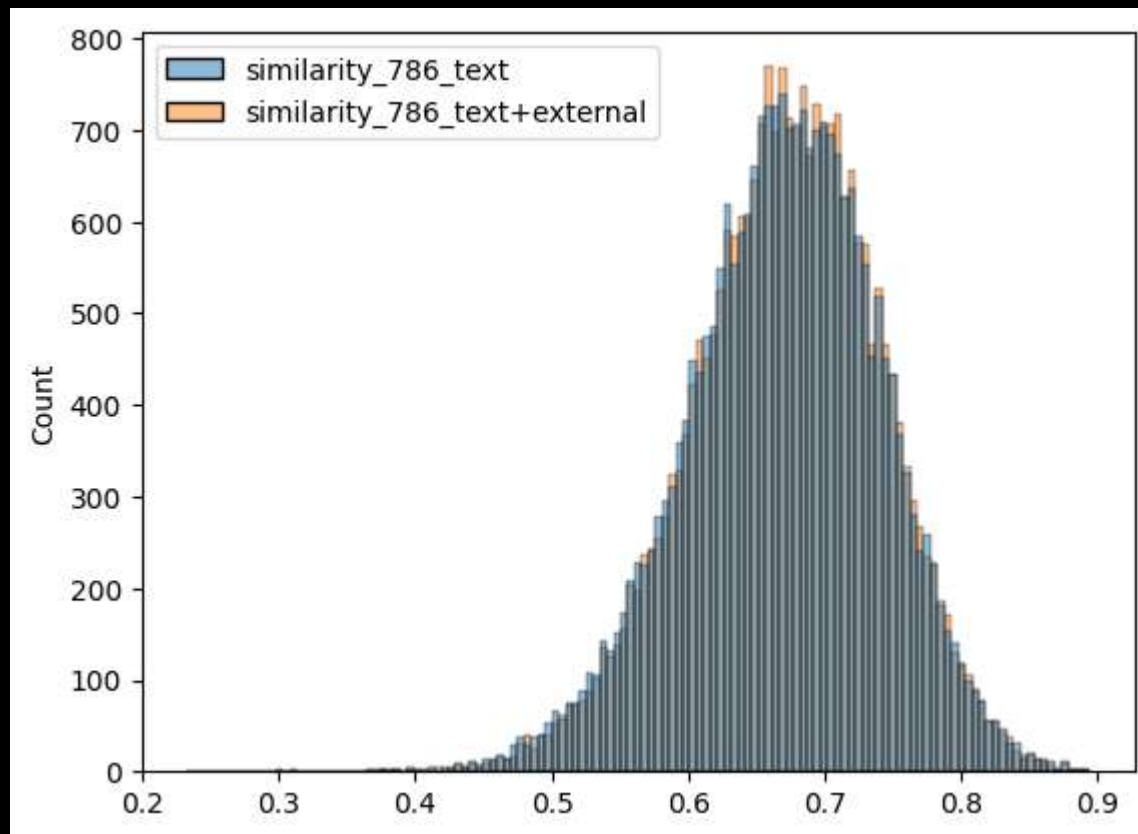


FROM NEWS TO INSIGHT

✂ Appendix

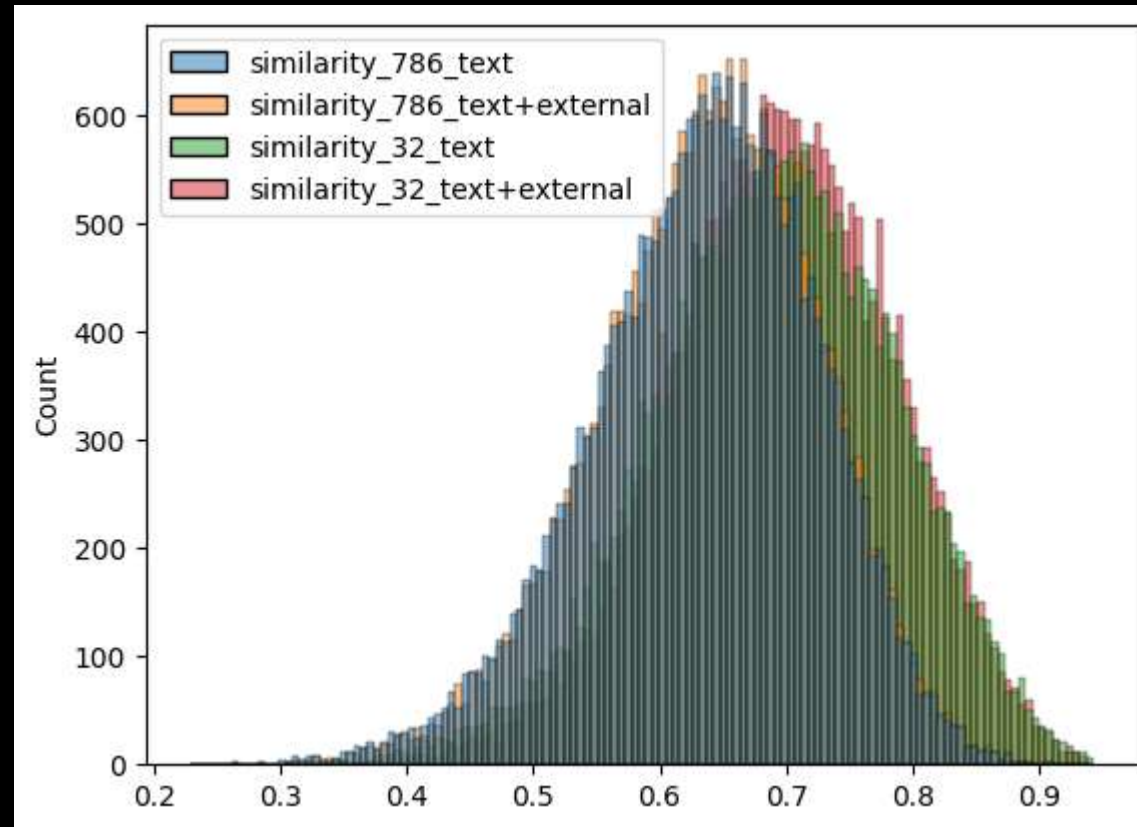
과거 유사 뉴스 모델 학습

타겟 유사도 정의



과거 유사 뉴스 모델 학습

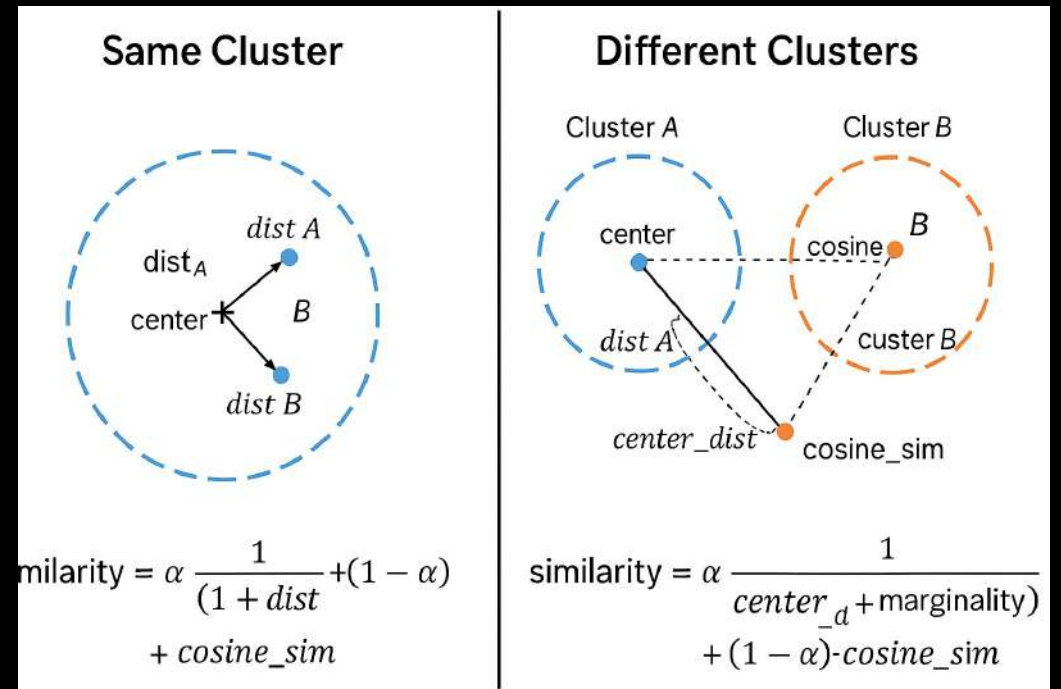
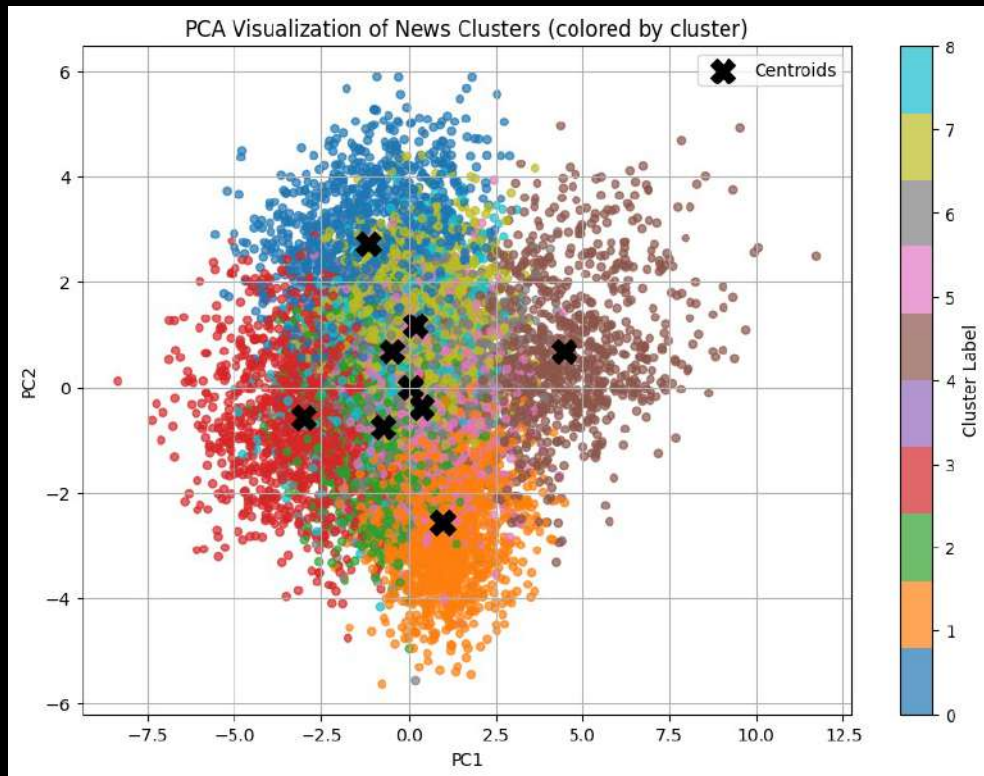
타겟 유사도 정의



Appendix

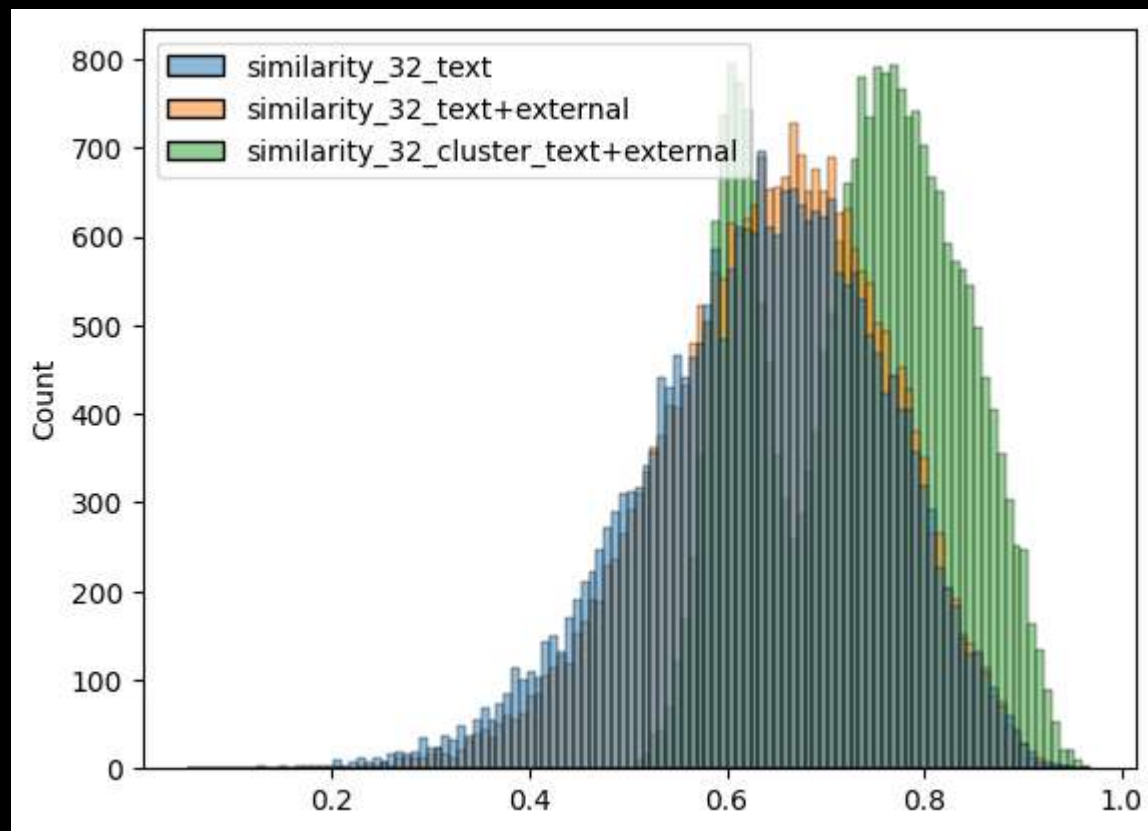
과거 유사 뉴스 모델 학습

타겟 유사도 정의



과거 유사 뉴스 모델 학습

타겟 유사도 정의



[챗봇 – 메시지 영속성 보장 관련]

현재 챗봇의 목적 : 미래 사건에 대한 과거 유사 뉴스를 찾아주는 챗봇

예상 사용자 100명 예상 &

과거 뉴스 DB에 영속적으로 보관하여 사용자가 재접속하여 같은 질문을 보내도 똑같은 결과로 답변 가능

→ 메시지 자체의 영속성이 필수는 아님

Appendix

[메시지 브로커 후보군]

Kafka

- 대규모 데이터 처리에 최적화되어 있어 초당 수천~수만 메시지를 안정적으로 처리 가능
- 메시지를 디스크에 영속적으로 저장하여 소비자가 구독 중이지 않아도 메시지 유실 없이 재처리가 가능함

→ 예상 사용자 100명 이내, 메시지 중복 소비나 재처리까지 요구되지 않아 과도한 스펙으로 판단함

Redis Stream

- Kafka보다 경량화된 구조로 운영 및 설정이 단순하며, Redis 환경에서 쉽게 구축 가능함
- 메시지를 로그 형태로 저장해 영속성을 제공

→ 예상 사용자수와 서비스 특성상 필요 이상의 복잡도가 발생, Stream의 영속성은 요구되지 않음

RabbitMQ

- 다양한 라우팅 전략(Direct, Topic, Fanout 등)을 지원해 메시지를 유연하게 분배 가능
- 메시지 전달의 신뢰성 보장

→ RabbitMQ의 다양한 라우팅 전략은 필요하지 않고 메시지 전달의 신뢰성 또한 중요하지 않음

Redis Pub/Sub

- 기존에 실시간 뉴스, 하이라이트 뉴스 캐싱에 사용하던 Redis 인프라를 그대로 활용 가능
- 단순 구조로 빠르고 실시간성에 최적화

→ 기존 인프라 재활용과 챗봇의 실시간 전송 요구에 가장 적합하다고 판단함

Appendix

여러 사용자 요청에 대한 응답 처리

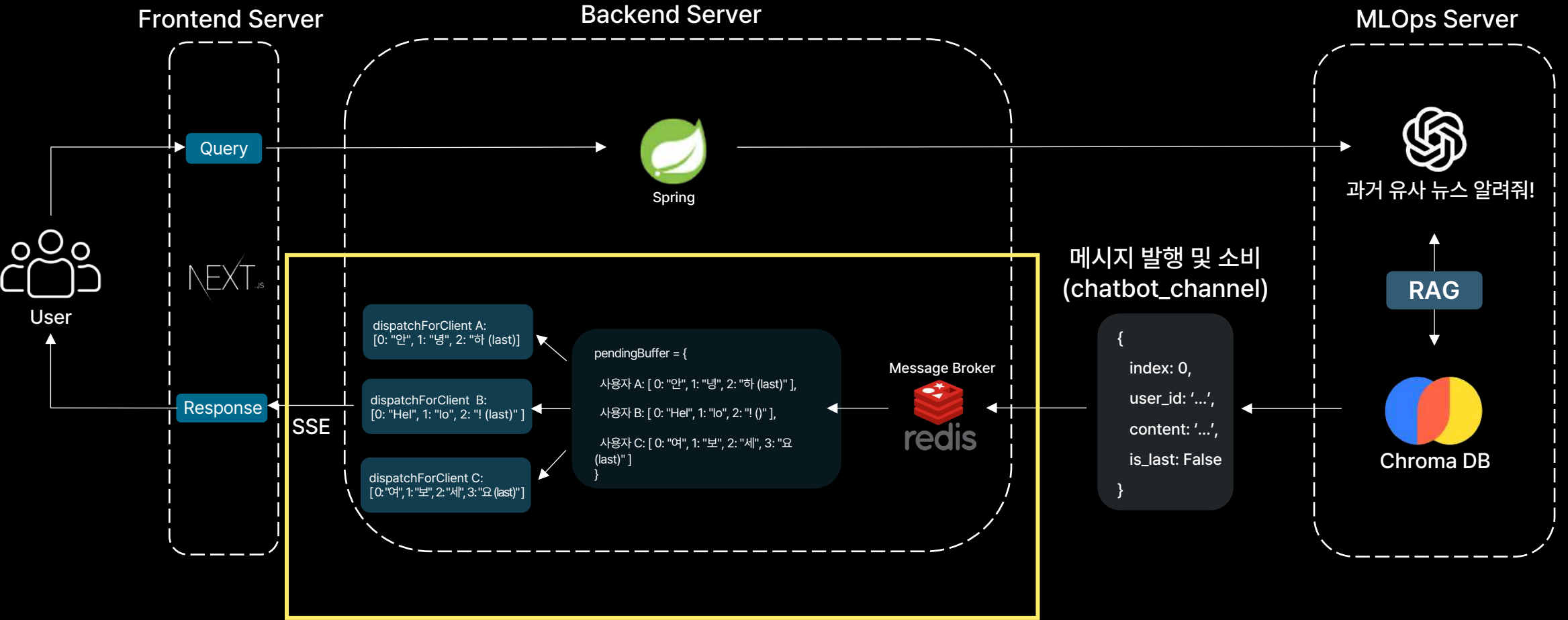
- Redis Pub/Sub은 단순한 채널 기반 브로드캐스트로, 수신자 구분 기능이 없음
- Redis는 단일 스레드로 메시지 순서를 보장하지만, Spring 서버의 Redis 구독 처리 워커는 병렬 처리되므로 응답 순서가 섞임

해결 방법

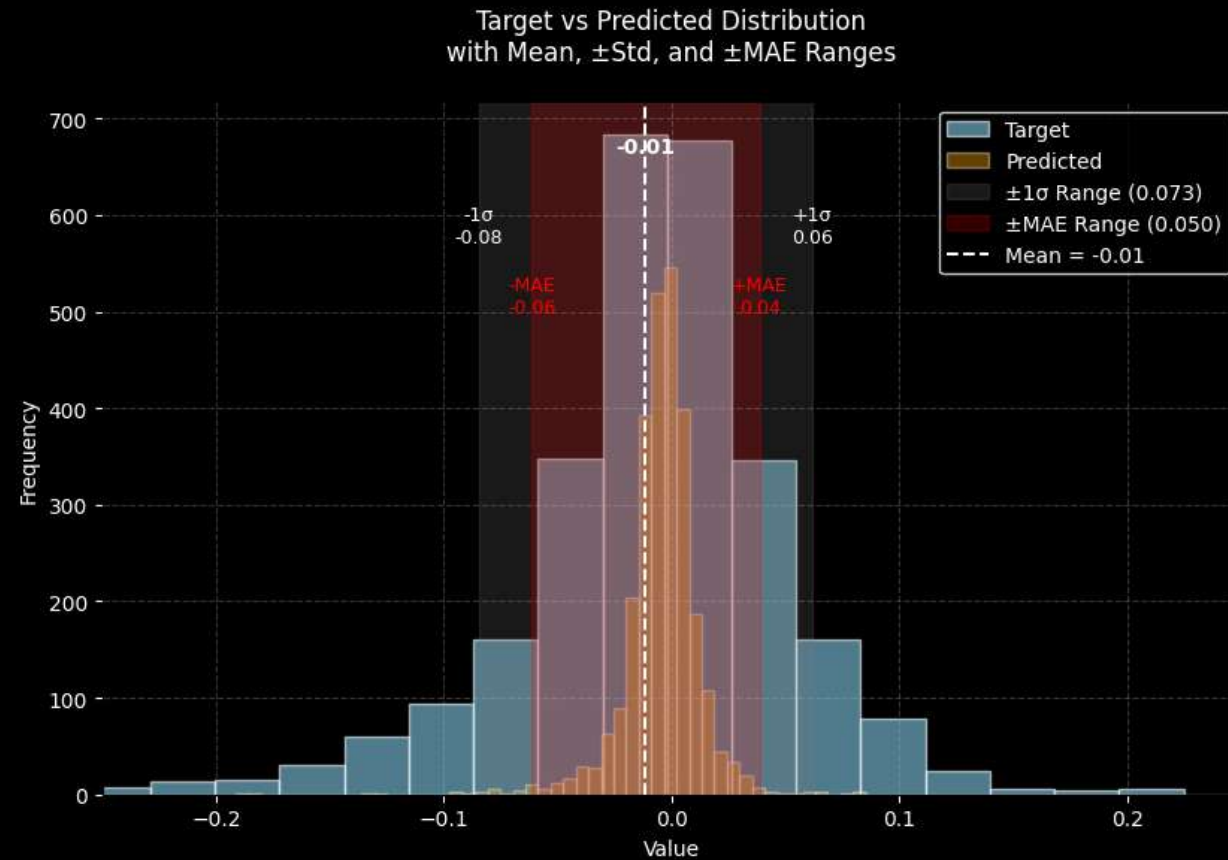
- 각 사용자마다 고유 client_id를 부여 + ML 응답에 Index 추가
- ConcurrentHashMap, ConcurrentSkipListMap를 사용해 pending buffer 저장소 구성 : 멀티쓰레드 + 순차 보장

- ▶ ConcurrentHashMap : HashMap처럼 키-값으로 데이터를 저장하지만 멀티쓰레드 환경에서도 안전
- ▶ ConcurrentSkipListMap : 여러 스레드가 동시에 삽입/삭제/탐색 작업을 안전하게 수행 가능, 키 순서대로 정렬된 상태를 유지

Appendix



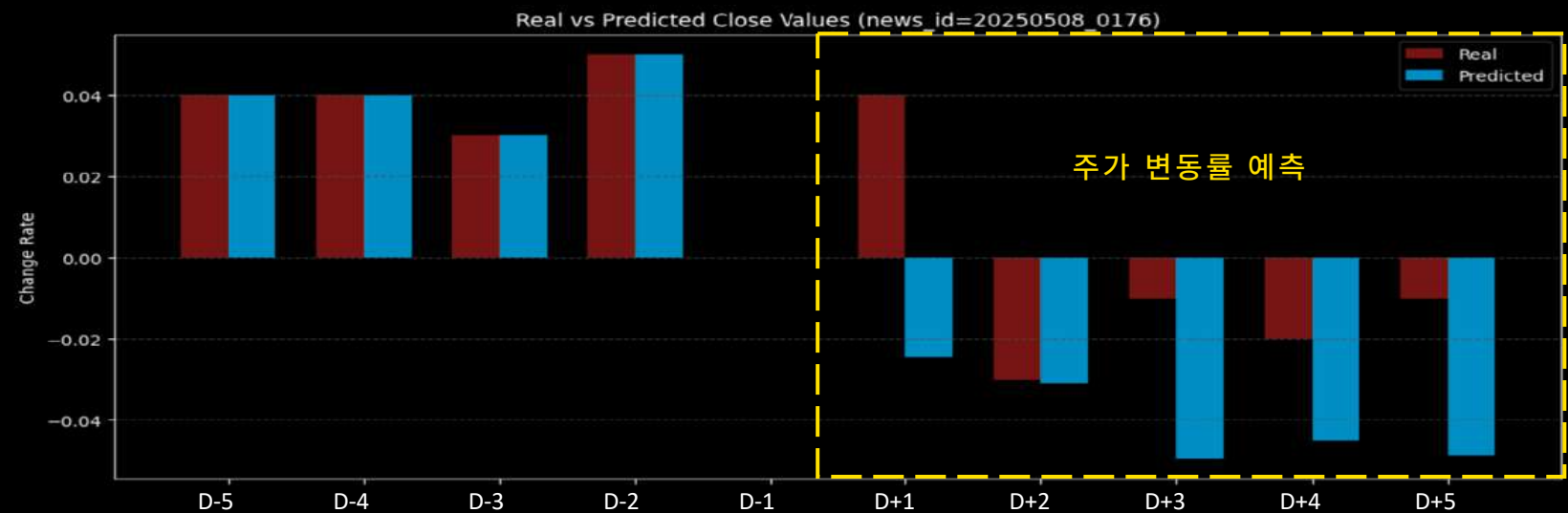
Appendix



MAE: 0.05

낮은 예측 신뢰도

Appendix



실제 최대 주가 변동률: 0.06

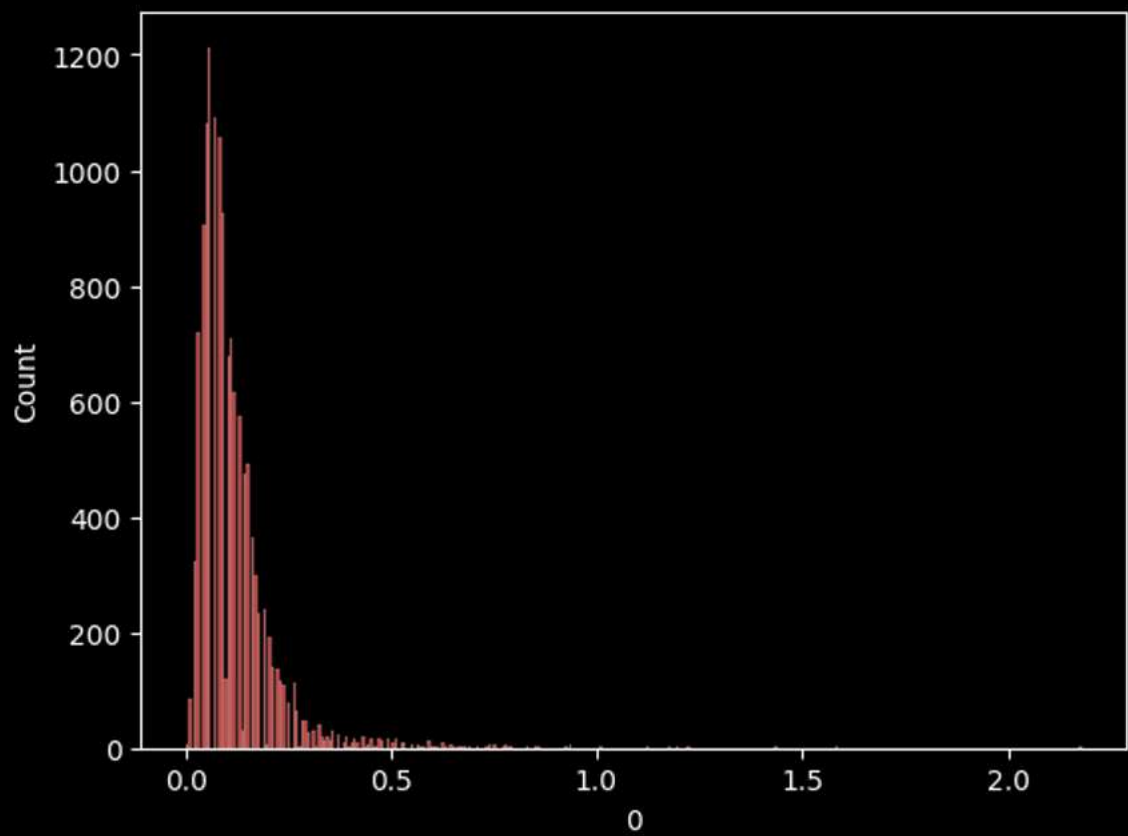
예측 최대 주가 변동률: 0.08



큰 차이 없음

뉴스 중요도 계산에는 모델 성능의 문제 없음을 판단

각 뉴스의 주가 최대 변동률 분포



count	13821.00000
mean	0.113848
std	0.096847
min	0.000000
25%	0.060000
50%	0.090000
75%	0.140000
max	2.180000