

**주식예측 해커톤**

**MLP LAB 임경태**

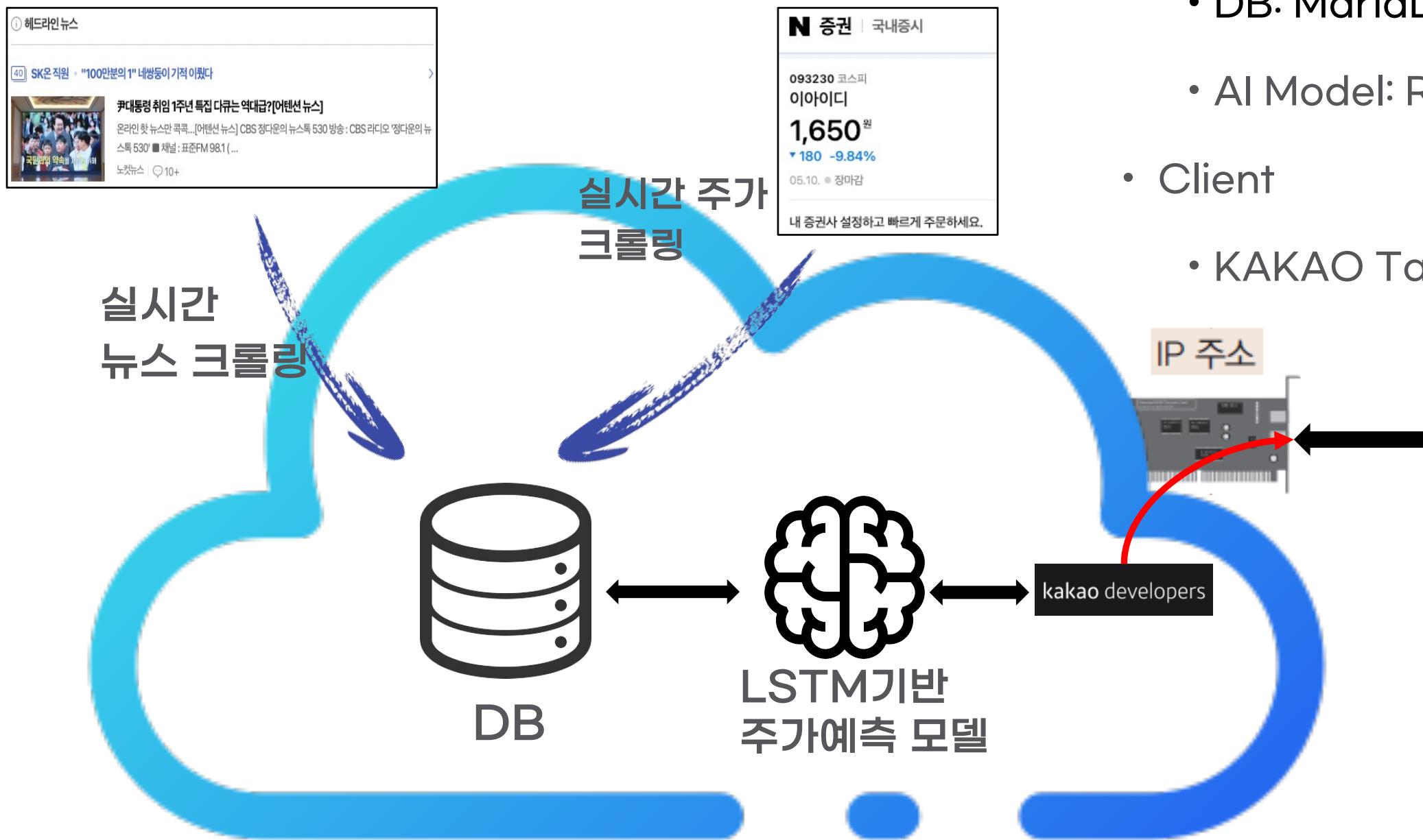


# 01 빅데이터 기술을 활용한 주가예측 해커톤!

- (목적) 실시간으로 주가를 예측하는 주가예측 어플리케이션을 구현한다. 총 3파티로 나누어 협업
  - **데이터전처리**: 뉴스/가격 데이터 **분단위** 크롤링
  - **빅데이터 저장**: Table 설계 후 sqlite에 데이터 실시간 저장 → sql을 통해 텍스트 데이터 분석 후 유의미한 키워드 도출
  - **모델학습**: LSTM, Transformer등 모델 학습 --> 도출된 feature를 토대로 학습
- (결과물) Github 프로젝트를 만들고 업로드

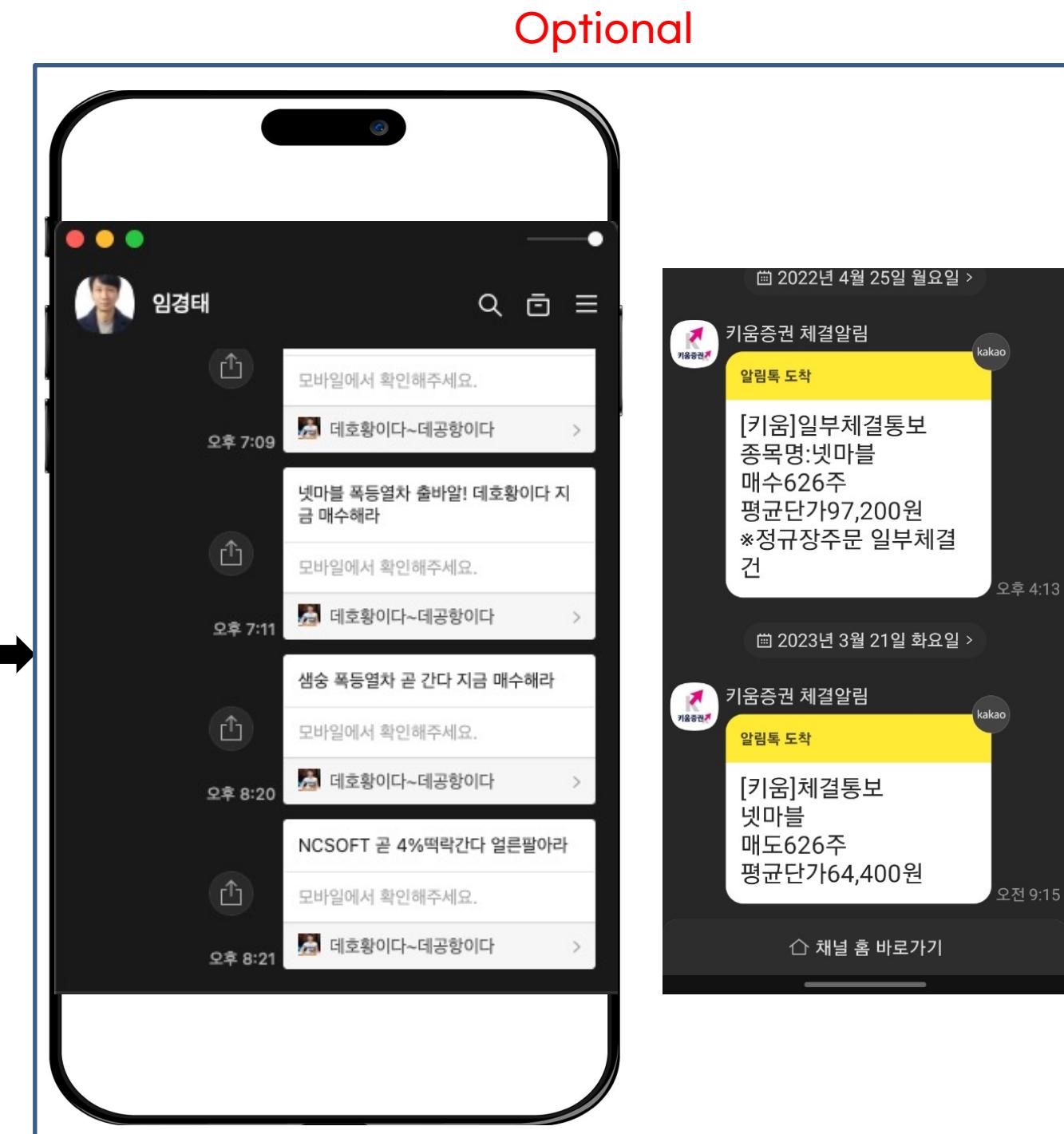
# 01 빅데이터 기술을 활용한 주가예측 해커톤!

- 제안하는 전체 시스템 구성



## 시스템 구성 정보

- Server-side
  - DB: MariaDB
  - AI Model: RNN
- Client
  - KAKAO Talk



## 02. 세부사항

# 02 데이터 수집 가이드라인

## • 데이터 수집 방법은 자유 다만 분단위 데이터를 수집!

- 특정 종목에 대한 가격 크롤링
- 특정 종목에 대해 네이버 증시 뉴스 크롤링 (<https://finance.naver.com/item/news.naver?code=005930>). 혹은 API 활용

The screenshot shows the Naver Finance website for Samsung Electronics (code: 005930). It displays the current price (77,600), high (78,200), low (77,700), and volume (5,250,002). Below this, there's a news feed section with various headlines related to Samsung Electronics.

The screenshot shows the NAVER Developers documentation for the News Search API. It includes sections for '블로그' (Blog), '뉴스' (News), '책' (Books), '성인 검색어 판별' (Adult search term identification), '백과사전' (Encyclopedia), '카페글' (Cafe posts), '지식iN' (Knowledge iN), '지역' (Region), '오타변환' (Orthographic conversion), '웹문서' (Web document), '이미지' (Image), '쇼핑' (Shopping), '전문자료' (Professional materials), and '참고 사항' (Reference). A code example for making a GET request to the API is provided:

```
> GET /v1/search/news.xml?query=%EC%A3%BC%EC%8B%9D&display=10&start=1&sort=sim HTTP/1.1
> Host: openapi.naver.com
> User-Agent: curl/7.49.1
> Accept: /*
> X-Naver-Client-Id: {애플리케이션 등록 시 발급받은 클라이언트 아이디 값}
> X-Naver-Client-Secret: {애플리케이션 등록 시 발급받은 클라이언트 시크릿 값}
>
```

Below this, there's a '요청 예' (Request example) section with a curl command:

```
curl "https://openapi.naver.com/v1/search/news.xml?query=%EC%A3%BC%EC%8B%9D&display=10&start=1&sort=sim" \
-H "X-Naver-Client-Id: {애플리케이션 등록 시 발급받은 클라이언트 아이디 값}" \
-H "X-Naver-Client-Secret: {애플리케이션 등록 시 발급받은 클라이언트 시크릿 값}" -v
```

## 03 빅데이터 저장

- **sqlite를 활용한 빅데이터 저장 및 분석**

- 저장방법: 클라우드, local, google drive 등 저장소는 자유
- 데이터저장: Table 설계 및 크롤링한 데이터 저장
- 데이터분석: sql을 이용해 워드클라우드, 단어 통계 등을 도출하고 해당 종목과 관련된 유의미한 키워드 분석결과 도출
- 데이터활용: 기계학습을 위한 데이터 전달

모델 개발담당자와 사전 조율해서 기계학습에 활용될 데이터의 포맷과 샘플 예제  
데이터를 미리 정해둬야 병렬처리가능

## 03 모델 개발

- 시계열 모델을 활용한 기계학습 모델링

- 도입모델: LSTM, Transformer, BERT, GPT등 자유
- 모델예측결과: 주가 혹은 상승/하락 레벨 (예, 1%, 5%)
- 성능평가: 성능평가 결과 어느정도 신뢰성이 있는지 분석
- 모델활용: 실시간으로 예측하도록 구현

## 04 프로젝트 관리

- 진행 완료한 프로젝트를 github에 업로드
  - 14주차 codes/github/README.MD 파일참조

## 05 만나서 반가웠고

- 팀프로젝트는 험난한 정치의 여정임
  - 근데 앞으로 대부분의 인생이 팀프로젝트임.. 이 강의가 도움이 되었길



**감사합니다.**