**회 의 록**

| **회의 일시** | 2025.05.01 | **회의 시간** | 14:30 ~ 16:00 | **작성자명** | 김효정 |
| --- | --- | --- | --- | --- | --- |
| **안건** | API 수집 방안 및 레이블 정의 | | | **장소** | 4층 휴게실 |

| **회의 내용 (요약)** |
| --- |
| **✅ API 수집 방안 및 레이블 정의에 대한 회의록****1. 안건**  * 뉴스 API 수집 방식 및 정확도 개선 * 레이블 정의  **1.2. 참석자별 의견 정리** ✅ **우성욱**   * 뉴스 데이터 수집 시 유료 API 활용 여부 논의 * CNN, CNBC는 뉴스 API로 수집 가능하며, 직접 크롤링보다 정확도가 높을 수 있음 * 디스크립션만으로 감성 분석 진행 중이며, 본문 수집이 어렵기 때문에 우선 사용 중   ✅ **이제환**   * NewsAPI 수집중. 하루에 수집량 제한 있음 * 디스크립션 만으로 감성분석이 되는지 의문 제기 * 유료 API 선택 시 Polygon외 대안 검토 필요 * 종가 기반 레이블 제안   ✅ **김나현**   * CNBC 수집 중 데이터 정확도 문제 발생 * 트루스소셜, X 등 SNS 기반 데이터 수집 필요성 제기 * 레이블은 전략보다 우선 설정되어야 한다고 의견 제시   ✅ **김효정**   * API 수집 시 많은 양의 데이터보다 정확도가 중요하다고 생각 * 수익률을 기준으로 한 범주형 레이블(강한 상승, 보류, 하락 등)도 고려 가능   ✅ **김현태**   * 정확도보다는 다양한 양의 데이터 수집이 필요하다는 의견 제시 * 단타 기반 예측 모델을 위해 1분, 5분, 10분 뒤 종가를 레이블로 활용할 것을 제안   ✅ **마민우**   * 이동 평균선 설정 시 배수를 활용한 분석 제안 * 레이블 기준은 최소 3분, 최대 15분 후 종가 예측으로 설정 제안   **2. 결론**   * 뉴스 API를 우선 활용하고, CNN과 CNBC는 MCP 방식을 이용해서 수집 시도 * SNS API 수집은 트루스소셜과 X(Twitter)를 중심으로 진행 예정   **5. 향후 논의 주제**   * 감성 분석용 콘텐츠 처리 방식 (본문 vs 디스크립션) * 각자 하고싶은 모델 선택 후 팀 나누기   **6. 다음 회의 일정: 2025-05-05(월)???** |

| **참석자** | **성명** | **불참자** | **사유** |
| --- | --- | --- | --- |
| 김나현 |  |
| 김지희 |  |
| 김현태 |  |
| 김효정 |  |
| 마민우 |  |
| 우성욱 |  |
| 이제환 |  |