|  |  |
| --- | --- |
| **팀명 : 도개교** | **일자 : 25/09/02** |

|  |  |  |
| --- | --- | --- |
| 팀원 이름(역할) | (09/02) 활동 내역 | 진행도 (전체 기준) |
| 김시온  (ㅇㅇ) | 원티드 직무별 스킬 크롤링, 잡코리아 공고별 지원자 정보(지원자 수, 학력, 연령, 성별),  데이터 기반 생성 기준 논의,  점핏 보고서 기반 구직자 직무,연차 비율 리스트 정리,  원티드 직무별 스킬 <-> 점핏 직무 기반으로 포맷 정리 | 15% |
| 향후 진행사항 : 직무별 스킬 공백란 채우기 (로드맵 사이트 기준), 데이터 생성 |
| 김지민  (팀원) | 테이블 명세서(hf\_meta) 필드 작성,  huggingface 모델 데이터 수집 코드 작성 완료(메타데이터, 파일데이터, 토론데이터),  databricks 공부(delta table, delta lake, unity catalog, dbfs, pyspark),  Workspace와 Unity Catalog에 데이터 적재 성공 | 20% |
| 향후 진행사항: 테이블 명세서(hf\_meta, hf\_files, hf\_community) 작성 완료, databricks에 싣기,  크롤링최적화 크롤링최적화 크롤링최적화 크롤링최적화 크롤링최적화 크롤링최적화 크롤링최적화 크롤링최적화  진정슨;; |
| 박형진  (팀원) | 점핏 크롤링 코드 작성 및 리뷰 | 15% |
| 향후 진행사항 : 크롤링 실행 및 수정, DB 적재 |
| 송누림  (팀원) | Databricks 클러스터 재생성 (최대 core 수 8개로 하향 조정), Databricks 공부(delta table, delta lake, unity catalog, dbfs, pyspark), 점핏 사이트 크롤링 코드 점검 및 병합 | 10% |
| 향후 진행사항 : 점핏 사이트 크롤링 후 csv, delta table, db 적재 |
| 이민지  (팀원) | 원티드, 잡코리아 크롤링 시도 (...), 더미 테이블 명세서 1차 작성, 더미 테이블 생성 시나리오 설계, 데이터 스케일링 비교 | 13% |
| 향후 진행사항 : 더미 테이블 연령/성별/학력 생성, 추가 feature 생성 후 테이블 완성 |

|  |  |
| --- | --- |
| **담당자 피드백** | 초기 기획을 명확히 하고, 테이블 설계까지 빠르게 진행중이시네요.  본격적인 크롤링 작업 시작 전에 필요한 정보들에 대해 충분히 검토하고 진행하시면 좋겠습니다.  프로젝트 초반에 에너지 소진되지 않도록 컨디션 조절 부탁드립니다. |
| **담당자 확인** | 임대건 |