**프로젝트 진행일지(멘토링)**

**[세상의 모든 AI]**

|  |  | **작성자** | 이재혁 |
| --- | --- | --- | --- |
|  |  | **날 짜** | 2022. 11. 10(1회) |
|  |  | **시 간** | 14시 00분 ~ 15시 30분 |
| **팀 명** | 코람페(코딩해서 람보르기니, 페라리 사자) | | |
| **팀 장** | 이재혁 | | |
| **팀 원** | 임철성, 정우현, 주성민, 이민흠, 손재형 | | |
| **멘 토** | (주)젠데이터\_함민혁, 최세아 | | |
| **멘 토 링 내 용** | | | |
| [진행사항 1]   * 가상 데이터 생성   [세부내용]   * 학습용 데이터 수량 3,000건을 목표로 생성 예정 * 데이터 생성 과정 참여 희망함을 전달 * 결과값(추천 패키지) 지정과 관련하여 규칙 필요   [진행사항 2]   * 정규 멘토링 시간 결정   [세부내용]   * 화요일, 목요일 14시 * 멘토사 외부 일정으로 변동 가능성 있음 * 일 정 변경 필요 경우 사전에 공지 및 조율 | | | |

| **멘 토 링 사 진** |
| --- |
|  |

**[세상의 모든 AI]**

|  |  | **작성자** | 이재혁 |
| --- | --- | --- | --- |
|  |  | **날 짜** | 2022. 11. 15(2회) |
|  |  | **시 간** | 14시 00분 ~ 15시 30분 |
| **팀 명** | 코람페(코딩해서 람보르기니, 페라리 사자) | | |
| **팀 장** | 이재혁 | | |
| **팀 원** | 임철성, 정우현, 주성민, 이민흠, 손재형 | | |
| **멘 토** | (주)젠데이터\_함민혁, 최세아 | | |
| **멘 토 링 내 용** | | | |
| [진행사항 1]   * Github Repository 공유   [세부내용]   * <https://github.com/gendata-git/semo-ai>   [진행사항 2]   * 가상 데이터 생성   [세부내용]   * 샘플데이터 전달 및 데이터 구조 설명 * 가상 데이터 생성 가이드라인 제공 * 추후에 가상 데이터 아닌 실제 고객 대상으로 수집한 정보 전달 예정 → 테스트셋으로 사용   [진행사항 3]   * 수급 요청 내용   [세부내용]   * 전달한 샘플 데이터에 빈 칸 채워서 35,000건 * cvs 또는 xlsx 파일로   [진행사항 3]   * 프론트 개발   [세부내용]   * 디자인 협업툴 결정   + XD 또는 제플린 * 개발 가능 여부 검토 | | | |

| **멘 토 링 사 진** |
| --- |
|  |

**[세상의 모든 AI]**

|  |  | **작성자** | 이재혁 |
| --- | --- | --- | --- |
|  |  | **날 짜** | 2022. 11. 17(3회) |
|  |  | **시 간** | 14시 00분 ~ 15시 30분 |
| **팀 명** | 코람페(코딩해서 람보르기니, 페라리 사자) | | |
| **팀 장** | 이재혁 | | |
| **팀 원** | 임철성, 정우현, 주성민, 이민흠, 손재형 | | |
| **멘 토** | (주)젠데이터\_함민혁, 최세아 | | |
| **멘 토 링 내 용** | | | |
| [진행사항 1]   * 다음주 회의 일정   [세부내용]   * 화, 금요일 오후 2:10   [진행사항 2]   * 가상 데이터 생성   [세부내용]   * 30,000건 생성 완료 * 숫자로 입력 받는 질문 추가 원함 * 숫자 데이터가 포함된 데이터셋 별도 생성하여 기존의 데이터셋과 학습 결과 비교   [진행사항 3]   * 프론트 개발 범위   [세부내용]   * 웹페이지 기본 구성 * 설문조사 양식 만들기 * 입력 안 된 내용 있으면 알림 팝업 띄우기   [진행사항 4]   * 태스크 관리 방법   [세부내용]   * 칸반보드 적극 활용 | | | |

| **멘 토 링 사 진** |
| --- |
|  |

**[세상의 모든 AI]**

|  |  | **작성자** | 이재혁 |
| --- | --- | --- | --- |
|  |  | **날 짜** | 2022. 11. 22(4회) |
|  |  | **시 간** | 14시 00분 ~ 15시 00분 |
| **팀 명** | 코람페(코딩해서 람보르기니, 페라리 사자) | | |
| **팀 장** | 이재혁 | | |
| **팀 원** | 임철성, 정우현, 주성민, 이민흠, 손재형 | | |
| **멘 토** | (주)젠데이터\_함민혁, 최세아 | | |
| **멘 토 링 내 용** | | | |
| [진행사항 1]   * 패키지 상품 수 축소(6종 >>> 3종)   [세부내용]   * 모델 정확성을 높이기 위해 패키지 상품 가짓수 축소 * 올인원, 데이터셋 구축, AI 개발   [진행사항 2]   * 데이터 구축 사례 제시 기능 추가가   [세부내용]   * 신규 기능 추가에 따른 화면 구성 고민 필요   1. AI 모델   2. 화면 구성   3. 업무 분담담 * 패키지 상품 선택 후 고객의 이해를 돕기 위해 데이터 구축 사례를 제시하는 기능 추가 * 고객이 AI 도입을 원하는 사업의 업종 및 분야를 자세하게 입력 받은 후 유사한 데이터 구축 사례 제시 * 사례는 공공데이터 사이트 웹크롤링 또는 오픈 API 사용 * - 참고 사이트 : AI Hub(<https://www.aihub.or.kr/devsport/openapireqst/reqst.do?currMenu=124&topMenu=101>) * - 참고 사이트 : 공공데이터 포털(<https://www.data.go.kr/index.do>) * 사례 제시 방법은 화면에 해당 내용을 띄워주거나 샘플 데이터 다운로드 지원원   [진행사항 3]   * 요구사항 정의서 작성   [세부내용]   * 신규 기능에 대한 요구사항 정의서 작성 * 화면 설계서 작성성 * 22.11.22(금)까지 | | | |

| **멘 토 링 사 진** |
| --- |
|  |

**[세상의 모든 AI]**

|  |  | **작성자** | 이재혁 |
| --- | --- | --- | --- |
|  |  | **날 짜** | 2022. 11. 25(5회) |
|  |  | **시 간** | 14시 00분 ~ 15시 00분 |
| **팀 명** | 코람페(코딩해서 람보르기니, 페라리 사자) | | |
| **팀 장** | 이재혁 | | |
| **팀 원** | 임철성, 정우현, 주성민, 이민흠, 손재형 | | |
| **멘 토** | (주)젠데이터\_함민혁, 최세아 | | |
| **멘 토 링 내 용** | | | |
| 샘플링 팀 업무 진행상황  1. AI 허브 내 384개 데이터셋들의 타이틀, 소개, 구축목적, 사용모델, 링크를 엑셀에 샘플링(완료)  2. 사업자등록시 사용하는 업종분류 대,중,소분류를 통계청 자료를 통해 조사후 엑셀에 정리(완료)  3. 업종분류중 소분류에 적합한 데이터셋 분류하기(75% 진행 중)  - A칼럼 소분류(업종명)  - B칼럼(업종과 관련도가 가장 높다고 판단되는 데이터셋 3개)  - C칼럼(그 외 n개 칼럼)  크롤링 팀 업무 진행상황   * AI허브 사이트 접속, 전체 데이터 제목 크롤링 및 정리 * 사용자가 입력한 내용 기준 유사도 높은 제목 3가지 선정 * 현재는 사용자 입력 내용 포함 여부로 조사한것이라 추후 word2vec등 유사도 분석방법으로 교체 필요 * 추출한 3가지 제목을 검색창에 입력하여 해당 페이지에 접속하여 데이터개요에 해당하는 소개 및 구축목적 크롤링 * 크롤링해온 데이터셋 제목, 소개, 구축목적 데이터프레임 형태로 제작 및 csv형태로 저장 * 해당 데이터셋 '어노테이션 포맷 및 데이터 구조'항목 스크린샷 및 png파일 형태로 저장 * 구축활용가이드 다운로드 링크주소 복사 * 데이터셋 제목, pdf파일 형태(데이터 설명서or구축활용가이드), 다운로드 링크 데이터프레임 형태로 제작 및 csv 형태로 저장 * pdf의 경우 구축활용가이드가 존재시 구축활용가이드만 가지고 오고 구축활용가이드가 없으면 데이터설명서를 다운로드하고 둘다 없는경우에는 pdf파일링크가 없다고 표시하는 형태   추후 계획  - 우현, 재혁 28(월)까지  ---한 개 분야별로 3개 데이터로 분리한 것을 웹페이지 상에서 선택해서 보여줄 수 있도록하기 =>AI 불필요  ---대/중/소분류에서 구성된 데이터 셋에서 어떤 것이 적절한지 보여줄 것 => 인터페이스 신경쓸 것  - 철성, 민흠, 재형, 성민  ---회사(고객)의 홈페이지 정보와 데이터를 매칭을 시킬 것인지 단어와 단어를 매칭을 시킬것인지 =>문제 해결  ---슬랙에 올라온 광인사 채용 공고 톡방에 올라온 회사 3~4 회사를 포함시켜 유사도 분석  ---코사인 유사도 보다는 word2vec을 쓰는게 좋아보임 => 크롤링팀(철성,민흠) + 재형, 성민  To - Do - List  1. 타이틀, 소개, 구축목적을 크롤링해서 샘플링 팀의 엑셀 파일 양식대로 맞출 것  2. 코사인유사도, word2vec, FastText 사용해 두 개의 텍스트 사이에 점수를 내는 방법 찾기(Plan A) => 재형, 성민  3. 웹페이지 관련 Plan C => 우현, 재혁  4. 다중레이블 방식은 Plan B이지만 아이디어 부족으로 보류  5. 각 영역별로 세부적인 계획 세우기  ---Flask 지식 필요--- | | | |

| **멘 토 링 사 진** |
| --- |
|  |

**[세상의 모든 AI]**

|  |  | **작성자** | 이재혁 |
| --- | --- | --- | --- |
|  |  | **날 짜** | 2022. 11. 29(6회) |
|  |  | **시 간** | 14시 00분 ~ 15시 00분 |
| **팀 명** | 코람페(코딩해서 람보르기니, 페라리 사자) | | |
| **팀 장** | 이재혁 | | |
| **팀 원** | 임철성, 정우현, 주성민, 이민흠, 손재형 | | |
| **멘 토** | (주)젠데이터\_함민혁, 최세아 | | |
| **멘 토 링 내 용** | | | |
| * 다음 회의 일정   + 22.12.01(목) 14:10 * 웹 크롤링 관련 문의사항   + 크롤링 속도(반복 주기)가 서버에 부담을 주지 않도록 고려   + 캡쳐하려는 element(어노테이션 포맷 및 데이터 구조)가 화면보다 큰 경우 잘리지 않는지 확인 * 소분류에 적합한 데이터셋 분류   + 사용자가 선택한 업종 소분류를 기준으로 전체 데이터셋에서 적합한 데이터셋을 한 번 필터링하여 리스트업 * 유사도가 모두 낮은 경우   + 기준을 높게(50%) 잡았을 때 에러 발생하여 낮게 조정   + 점수가 낮아도 항상 3개 게시물 추천 * 요청사항   + 구현 범위와 일정   + 성능 개선을 위한 계획   + 서비스 개념도 \*간단하게 | | | |

| **멘 토 링 사 진** |
| --- |
|  |

**[세상의 모든 AI]**

|  |  | **작성자** | 이재혁 |
| --- | --- | --- | --- |
|  |  | **날 짜** | 2022. 12. 01(7회) |
|  |  | **시 간** | 14시 00분 ~ 15시 00분 |
| **팀 명** | 코람페(코딩해서 람보르기니, 페라리 사자) | | |
| **팀 장** | 이재혁 | | |
| **팀 원** | 임철성, 정우현, 주성민, 이민흠, 손재형 | | |
| **멘 토** | (주)젠데이터\_함민혁, 최세아 | | |
| **멘 토 링 내 용** | | | |
| * 다음 회의 일정   + 22.12.06(화) 14:10 * 성능 개선 보고서   + 추후 성능 개선 보고서 작성을 위해 구현 당시의 성능 측정 및 기록 필요   + 이후 성능 개선 작업 진행 후와 비교하여 문서 작성   + 개선 작업 후에 수치 상에서 유의미한 변화가 없다면 코드 단순화 방법 위주로 작성   + 프론트 페이지 랜딩 속도   + 모델 성능 측정   + 모델 속도 측정 * 백로그 기록 (필수X, 장려O)   + 업무 상세하게 작성   + 개발 과정에 대한 히스토리 * 라이브러리 추천   + Fastai   + 플라스크와 유사 | | | |

| **멘 토 링 사 진** |
| --- |
|  |

**[세상의 모든 AI]**

|  |  | **작성자** | 이재혁 |
| --- | --- | --- | --- |
|  |  | **날 짜** | 2022. 12. 06(8회) |
|  |  | **시 간** | 14시 00분 ~ 16시 00분 |
| **팀 명** | 코람페(코딩해서 람보르기니, 페라리 사자) | | |
| **팀 장** | 이재혁 | | |
| **팀 원** | 임철성, 정우현, 주성민, 이민흠, 손재형 | | |
| **멘 토** | (주)젠데이터\_함민혁, 최세아 | | |
| **멘 토 링 내 용** | | | |
| * 다음 회의 일정   1. 22.12.09(금) 14:10 * 정확도, 속도 * 프로세스    1. 팀원소개   2. 모델 선택(특징 비교 안내문)   3. 회사명 입력      + AI-Hub에서 제목 크롤링      + 회사명 커리어에 검색하여 주요사업내용 저장      + 커리어에 검색 결과 없으면 주요사업내용 직접 입력 | | | |

| **멘 토 링 사 진** |
| --- |
|  |

**[세상의 모든 AI]**

|  |  | **작성자** | 이재혁 |
| --- | --- | --- | --- |
|  |  | **날 짜** | 2022. 12. 06(9회) |
|  |  | **시 간** | 14시 00분 ~ 16시 00분 |
| **팀 명** | 코람페(코딩해서 람보르기니, 페라리 사자) | | |
| **팀 장** | 이재혁 | | |
| **팀 원** | 임철성, 정우현, 주성민, 이민흠, 손재형 | | |
| **멘 토** | (주)젠데이터\_함민혁, 최세아 | | |
| **멘 토 링 내 용** | | | |
| * 다음 회의 일정    + 22.12.13(화) 14:10 * 백엔드 이슈(GET POST 통신) 대부분 해결 * html 테스트 진행 중 * 발표용 PPT 작성 중이며, 초안 완성되면 노션 업로드 예정 | | | |

| **멘 토 링 사 진** |
| --- |
|  |

**[세상의 모든 AI]**

|  |  | **작성자** | 이재혁 |
| --- | --- | --- | --- |
|  |  | **날 짜** | 2022. 12. 13(10회) |
|  |  | **시 간** | 14시 00분 ~ 16시 00분 |
| **팀 명** | 코람페(코딩해서 람보르기니, 페라리 사자) | | |
| **팀 장** | 이재혁 | | |
| **팀 원** | 임철성, 정우현, 주성민, 이민흠, 손재형 | | |
| **멘 토** | (주)젠데이터\_함민혁, 최세아 | | |
| **멘 토 링 내 용** | | | |
| * 다음 회의 일정    + 마지막 회의 * 발표 피피티 자료, 대본 컨펌 * html 배경 수정 * 멘토링 마무리 | | | |

| **멘 토 링 사 진** |
| --- |
|  |