**프로젝트 진행일지**

**[세모 AI]**

|  |  | | **작성자** | 이재혁 | |
| --- | --- | --- | --- | --- | --- |
|  |  | | **날 짜** | 2022. 11. 07~09. (1회) | |
| **팀 명** | 코람페(코딩해서 람보르기니, 페라리 사자) | | | | |
| **팀구성원** | **이름** | **역할** | | | |
| 이재혁 | 팀장(프로젝트 총괄 및 문서작성) | | | |
| 임철성 | 데이터 관리 및 모델링 | | | |
| 손재형 | 데이터 전처리 및 모델링 | | | |
| 이민흠 | 데이터 전처리 및 모델링 | | | |
| 주성민 | 데이터 전처리 및 모델링 | | | |
| 정우현 | 모델링, 웹 화면 구현 | | | |
| **일별 진행상황** | | | | | |
| **이름** | **업 무 내 용** | | | | **특이사항** |
| 이재혁 | 프로젝트 방향성 회의진행 | | | |  |
| 임철성 | 프로젝트 방향성 아이디어 제시 | | | |  |
| 손재형 | 프로젝트 방향성 아이디어 제시 | | | |  |
| 이민흠 | 프로젝트 방향성 아이디어 제시 | | | |  |
| 주성민 | 프로젝트 방향성 아이디어 제시 | | | |  |
| 정우현 | 프로젝트 방향성 아이디어 제시 | | | |  |

| **회 의 내 용** |
| --- |
| [안건 1]   * 프로젝트 방향성에 대한 회의   [세부내용]  1.기업으로부터 제공 받는 데이터에 대한 정보  2.프로젝트 주제에 대한 의문점 토의   * 토의 결과 멘토분들과의 2차 멘토링이 시급 * 2차 멘토링시 기업에서 제공할 데이터의 형태와 주제의 명확성 요청 드릴 예정   3.멘토링 기업과 컨택   * 데이터관련 및 산출문서 등에 관한 답변 수신 * 멘토링 기업으로부터 샘플 데이터 및 참고문서 수신 |

**[세모 AI]**

|  |  | | **작성자** | 이재혁 | |
| --- | --- | --- | --- | --- | --- |
|  |  | | **날 짜** | 2022. 11. 10 . ( 2회) | |
| **팀 명** | 코람페(코딩해서 람보르기니, 페라리 사자) | | | | |
| **팀구성원** | **이름** | **역할** | | | |
| 이재혁 | 팀장(프로젝트 총괄 및 문서작성) | | | |
| 임철성 | 데이터 관리 및 모델링 | | | |
| 손재형 | 데이터 전처리 및 모델링 | | | |
| 이민흠 | 데이터 전처리 및 모델링 | | | |
| 주성민 | 데이터 전처리 및 모델링 | | | |
| 정우현 | 모델링, 웹 화면 구현 | | | |
| **일별 진행상황** | | | | | |
| **이름** | **업 무 내 용** | | | | **특이사항** |
| 이재혁 | 프로젝트 방향성 회의진행 | | | |  |
| 임철성 | 프로젝트 방향성 아이디어 제시 | | | |  |
| 손재형 | 프로젝트 방향성 아이디어 제시 | | | |  |
| 이민흠 | 프로젝트 방향성 아이디어 제시 | | | |  |
| 주성민 | 프로젝트 방향성 아이디어 제시 | | | |  |
| 정우현 | 프로젝트 방향성 아이디어 제시 | | | |  |

| **회 의 내 용** |
| --- |
| [안건 1]   * 멘토분들과의 회의 진행   [세부내용]  1. 가상데이터 생성 건   * 멘토사에서 학습용 데이터 수량 총 3,000건을 목표로 생성 예정 * 가상데이터 생성 과정에 함께 참여하여 진행하기로함 * 결과값(추천 패키지) 지정과 관련하여 규칙 필요 * 규칙 정의 되면 멘토사로부터 세부내용 전달 받기로 함   2. 멘토링 시간 결정 건   * 화요일, 목요일 14시 * 멘토사 외부 일정 때문에 변동 가능성 있음 * 일정 변경 필요한 경우 사전에 공지 및 조율하기로 함 |

**[세모 AI]**

|  |  | | **작성자** | 이재혁 | |
| --- | --- | --- | --- | --- | --- |
|  |  | | **날 짜** | 2022. 11. 11 . ( 3회) | |
| **팀 명** | 코람페(코딩해서 람보르기니, 페라리 사자) | | | | |
| **팀구성원** | **이름** | **역할** | | | |
| 이재혁 | 팀장(프로젝트 총괄 및 문서작성) | | | |
| 임철성 | 데이터 관리 및 모델링 | | | |
| 손재형 | 데이터 전처리 및 모델링 | | | |
| 이민흠 | 데이터 전처리 및 모델링 | | | |
| 주성민 | 데이터 전처리 및 모델링 | | | |
| 정우현 | 모델링, 웹 화면 구현 | | | |
| **일별 진행상황** | | | | | |
| **이름** | **업 무 내 용** | | | | **특이사항** |
| 이재혁 | 프로젝트 회의진행, 기획서 작성 | | | |  |
| 임철성 | 기획서 작성 및 수정 | | | |  |
| 손재형 | 기획서 작성 및 수정 | | | |  |
| 이민흠 | 기획서 작성 및 수정 | | | |  |
| 주성민 | 기획서 작성 및 수정 | | | |  |
| 정우현 | 기획서 작성 및 수정 | | | |  |

| **회 의 내 용** |
| --- |
| [안건 1]   * 프로젝트 기획서 최종 검토   [세부내용]  1. 프로젝트 기획서 내용에 관한 회의   * 제안배경 검토 * 수요분석 검토 * 달성목표 및 달성전략 검토 * SWOT 분석 내용 정리 * 발전 방향 검토 후 결과 도출   + 웹 페이지 뿐만 아니라 앱 서비스 형태도 고려 |

**[세모 AI]**

|  |  | | **작성자** | 이재혁 | |
| --- | --- | --- | --- | --- | --- |
|  |  | | **날 짜** | 2022. 11. 14 . ( 4회) | |
| **팀 명** | 코람페(코딩해서 람보르기니, 페라리 사자) | | | | |
| **팀구성원** | **이름** | **역할** | | | |
| 이재혁 | 팀장(프로젝트 총괄 및 문서작성) | | | |
| 임철성 | 데이터 관리 및 모델링 | | | |
| 손재형 | 데이터 전처리 및 모델링 | | | |
| 이민흠 | 데이터 전처리 및 모델링 | | | |
| 주성민 | 데이터 전처리 및 모델링 | | | |
| 정우현 | 모델링, 웹 화면 구현 | | | |
| **일별 진행상황** | | | | | |
| **이름** | **업 무 내 용** | | | | **특이사항** |
| 이재혁 | 프로젝트 회의진행, 기획서 작성 | | | |  |
| 임철성 | 기획서 작성 및 수정, 기획서 발표 | | | |  |
| 손재형 | 기획서 작성 및 수정 | | | |  |
| 이민흠 | 기획서 작성 및 수정 | | | |  |
| 주성민 | 기획서 작성 및 수정 | | | |  |
| 정우현 | 기획서 작성 및 수정 | | | |  |

| **회 의 내 용** |
| --- |
| [안건 1]   * 프로젝트 기획서 최종 검토 * 기획 발표회   [세부내용]  1. 프로젝트 기획서 발표전 최종 점검   * 제안배경 점검 * 수요분석 점검 * 달성목표 및 달성전략 점검 * SWOT 분석 내용 점검 * 발전 방향 검토 후 결과 도출 점검 |

**[세모 AI]**

|  |  | | **작성자** | 이재혁 | |
| --- | --- | --- | --- | --- | --- |
|  |  | | **날 짜** | 2022. 11. 16. ( 5회) | |
| **팀 명** | 코람페(코딩해서 람보르기니, 페라리 사자) | | | | |
| **팀구성원** | **이름** | **역할** | | | |
| 이재혁 | 팀장(프로젝트 총괄 및 문서작성) | | | |
| 임철성 | 데이터 관리 및 모델링 | | | |
| 손재형 | 데이터 전처리 및 모델링 | | | |
| 이민흠 | 데이터 전처리 및 모델링 | | | |
| 주성민 | 데이터 전처리 및 모델링 | | | |
| 정우현 | 모델링, 웹 화면 구현 | | | |
| **일별 진행상황** | | | | | |
| **이름** | **업 무 내 용** | | | | **특이사항** |
| 이재혁 | 프로젝트 회의진행, 문서작성, 가상데이터 생성 | | | |  |
| 임철성 | 가상데이터 생성, 데이터 전처리, 모델 샘플링 | | | |  |
| 손재형 | 가상데이터 생성, 데이터 전처리 | | | |  |
| 이민흠 | 가상데이터 생성, 데이터 전처리, 모델 샘플링 | | | |  |
| 주성민 | 가상데이터 생성, 데이터 전처리, 모델 샘플링 | | | |  |
| 정우현 | 가상데이터 생성, 데이터 전처리 | | | |  |

| **회 의 내 용** |
| --- |
| [안건 1]   * 데이터 전처리 * 가상데이터 생성   [세부내용]  1. 데이터 전처리   * 데이터 칼럼 내용 숙지 * 서비스 대상의 AI 모델 여부에 따른 패키지 종류 판별 * 사업체의 종류 숙지 * 사업체 종류에 따른 AI 도입을 원하는 분야 판별   2. 가상데이터 생성   * 기업 또는 단체명 칼럼 생성 * 사업체 종류 칼럼 생성 * AI 기술 관련도 칼럼 생성 * 기술 도입 분야 칼럼 생성 * 데이터 보유 유무 칼럼 생성 * 모델 보유 유무 칼럼 생성 * 외주 작업 경험 유무 칼럼 생성 * 개선을 기대하는 어려움 칼럼 생성   + 위 칼럼들에 따른 추천 패키지 칼럼 생성   3. 가상데이터 함수화   * 추후 추가 데이터가 필요할 경우 또는 생성할 경우를 대비해(현재 3만여건), 필요할 데이터 수를 입력하면 자동으로 입력한 수만큼 데이터를 생성하도록 함수화   4. 샘플 모델링   * 생성된 가상데이터를 토대로 독립변수, 종속변수를 지정하여 모델링 진행(샘플 모델이기 때문에 추후 수정할 예정) |

**[세모 AI]**

|  |  | | **작성자** | 이재혁 | |
| --- | --- | --- | --- | --- | --- |
|  |  | | **날 짜** | 2022. 11. 18. ( 6회) | |
| **팀 명** | 코람페(코딩해서 람보르기니, 페라리 사자) | | | | |
| **팀구성원** | **이름** | **역할** | | | |
| 이재혁 | 팀장(프로젝트 총괄 및 문서작성) | | | |
| 임철성 | 데이터 관리 및 모델링 | | | |
| 손재형 | 데이터 전처리 및 모델링 | | | |
| 이민흠 | 데이터 전처리 및 모델링 | | | |
| 주성민 | 데이터 전처리 및 모델링 | | | |
| 정우현 | 모델링, 웹 화면 구현 | | | |
| **일별 진행상황** | | | | | |
| **이름** | **업 무 내 용** | | | | **특이사항** |
| 이재혁 | 프로젝트 회의진행, 문서작성, 데이터 관리 | | | |  |
| 임철성 | 샘플 모델 연구(word2vec) | | | |  |
| 손재형 | 샘플 모델 연구(word2vec) | | | |  |
| 이민흠 | 샘플 모델 연구(word2vec) | | | |  |
| 주성민 | 샘플 모델 연구(word2vec) 및 프론트 구상 | | | |  |
| 정우현 | 샘플 모델 연구(word2vec) 및 프론트 구상 | | | |  |

| **회 의 내 용** |
| --- |
| [안건 1]   * word2vec을 활용한 모델 선정 * 수업시간에 사용하였던 모델 중 이번 프로젝트에 사용할 수 있는 모델 알아보기   [세부내용]  1. word2vec을 활용한 모델 선정   * word2vec을 활용하여 단어 임베딩 모델 * ELMo를 활용하여 모델링   2. 강의안에서 사용 가능한 모델 찾아보기   * CBOW 또는 Skip-gram, LSTM, GRU, RNN * 순환신경망 사용해보기 * BERT도 사용해볼 수 있을지 의논 |

**[세모 AI]**

|  |  | | **작성자** | 이재혁 | |
| --- | --- | --- | --- | --- | --- |
|  |  | | **날 짜** | 2022. 11. 21. ( 7회) | |
| **팀 명** | 코람페(코딩해서 람보르기니, 페라리 사자) | | | | |
| **팀구성원** | **이름** | **역할** | | | |
| 이재혁 | 팀장(프로젝트 총괄 및 문서작성) | | | |
| 임철성 | 데이터 관리 및 모델링 | | | |
| 손재형 | 데이터 전처리 및 모델링 | | | |
| 이민흠 | 데이터 전처리 및 모델링 | | | |
| 주성민 | 데이터 전처리 및 모델링 | | | |
| 정우현 | 모델링, 웹 화면 구현 | | | |
| **일별 진행상황** | | | | | |
| **이름** | **업 무 내 용** | | | | **특이사항** |
| 이재혁 | 프로젝트 회의진행, 문서작성, 데이터 관리 | | | |  |
| 임철성 | 샘플 모델 연구(word2vec) | | | |  |
| 손재형 | 샘플 모델 연구(word2vec) | | | |  |
| 이민흠 | 샘플 모델 연구(word2vec) | | | |  |
| 주성민 | 샘플 모델 연구(word2vec) 및 프론트 구상 | | | |  |
| 정우현 | 샘플 모델 연구(word2vec) 및 프론트 구상 | | | |  |

| **회 의 내 용** |
| --- |
| [안건 1]   * 7번째 열('컨설팅으로 개선을 기대하는 어려움이 있나요?')만 영향을 끼치게 가상데이터를 생성함 * 7번째 열에서 랜덤 초이스로 패키지 결정하는 현행 시스템 * 컬럼을 모두 합쳐서 한 컬럼으로 만들어서 Word2Vec 활용 종속변수 예측   [세부내용]  1. 데이터 셋 y값은 세가지로 (올인원, 데이터셋 구축, ai모델 개발)  2. 데이터 셋 구축의 세부 항목/ 데이터셋 구축 패키지 선택 시 세부 내용  (추가 설문지, 설문 내용에 따른 예시 데이터 셋)  3. 분류 항목 별 예시 ai 모델 결과 값에 대한 시각화  4. 최종 ai 모델 추천  5. 다양한 형태(이미지 등)의 입력 데이터 추가 방안  6. 질문의 종류가 다양해야 함  7. 하나의 컬럼만 종속변수에 영향을 끼치는 것이 문제  [안건 2] AIHub 내 샘플을 3가지 정도 보여주는 방안   * 7번째 열('컨설팅으로 개선을 기대하는 어려움이 있나요?')만 영향을 끼치게 가상데이터를 생성함 * 7번째 열에서 랜덤 초이스로 패키지 결정하는 현행 시스템 * 컬럼을 모두 합쳐서 한 컬럼으로 만들어서 Word2Vec 활용 종속변수 예측   [세부내용]  1. 데이터 셋 y값은 세가지로 (올인원, 데이터셋 구축, ai모델 개발) |

**[세모 AI]**

|  |  | | **작성자** | 주성민 | |
| --- | --- | --- | --- | --- | --- |
|  |  | | **날 짜** | 2022. 11. 23. ( 8회) | |
| **팀 명** | 코람페(코딩해서 람보르기니, 페라리 사자) | | | | |
| **팀구성원** | **이름** | **역할** | | | |
| 이재혁 | 팀장(프로젝트 총괄 및 문서작성) | | | |
| 임철성 | 데이터 관리 및 모델링 | | | |
| 손재형 | 데이터 전처리 및 모델링 | | | |
| 이민흠 | 데이터 전처리 및 모델링 | | | |
| 주성민 | 데이터 전처리 및 모델링 | | | |
| 정우현 | 모델링, 웹 화면 구현 | | | |
| **일별 진행상황** | | | | | |
| **이름** | **업 무 내 용** | | | | **특이사항** |
| 이재혁 | 프로젝트 회의진행, 문서작성, 데이터 관리 | | | |  |
| 임철성 | 샘플 모델 연구(word2vec) | | | |  |
| 손재형 | 샘플 모델 연구(word2vec) | | | |  |
| 이민흠 | 샘플 모델 연구(word2vec) | | | |  |
| 주성민 | 샘플 모델 연구(word2vec) 및 프론트 구상 | | | |  |
| 정우현 | 샘플 모델 연구(word2vec) 및 프론트 구상 | | | |  |

| **회 의 내 용** |
| --- |
| [안건 1 - 다중 레이블 분류 작업(지도학습)]  - 소분류가 아닌 중분류  - 설문조사를 거치지 않고 텍스트 내용을 보고 자동으로 연결 시켜주는 방법  - 홈페이지 크롤링 필요  - 케라스 창시자가 알려주는 딥러닝 관련 책 참고 필요  [안건 2 - 모델 관련]  [세부내용]  - Word2Vec / FastText 활용(비지도학습)해서 단어 비교해서 유사도가 가장 높은 데이터 연결  - 세부내용 팀 결정 필요  - 기업 분류마다 3개씩 연결할 때 우리가 다중 레이블 분류도 하니깐 염두 필요  => 사진 이미지를 주면 강아지, 도마뱀을 맞추듯 구분되어야 함, AIHub 내 유사한  데이터 참조 필요  - 첫 번째 - 제일 관련 / 두 번째 - 약간 관련 / 세 번째 - 안씀 => 같은 컬럼 내에선 동일한  관련도  - 예시) 자동차 고장 관련 이미지 다수 => 똑같은 5개에서 어떻게 3개로 줄일 것인가?  - 무조건 가장 연관도 높은 것 3개를 찾는다. / 나머지 2개는 두 번째 컬럼에 넣는 식으로  [안건 3 - 샘플데이터 관련]  - Word2Vec 활용  - 소개 부분 어디서 발췌할 것인가 => 기준 설정 후 유사도 산출 필요 |

**[세모 AI]**

|  |  | | **작성자** | 주성민 | |
| --- | --- | --- | --- | --- | --- |
|  |  | | **날 짜** | 2022. 11. 24. ( 9회) | |
| **팀 명** | 코람페(코딩해서 람보르기니, 페라리 사자) | | | | |
| **팀구성원** | **이름** | **역할** | | | |
| 이재혁 | 팀장(프로젝트 총괄 및 문서작성) | | | |
| 임철성 | 데이터 관리 및 모델링 | | | |
| 손재형 | 데이터 전처리 및 모델링 | | | |
| 이민흠 | 데이터 전처리 및 모델링 | | | |
| 주성민 | 데이터 전처리 및 모델링 | | | |
| 정우현 | 모델링, 웹 화면 구현 | | | |
| **일별 진행상황** | | | | | |
| **이름** | **업 무 내 용** | | | | **특이사항** |
| 이재혁 | 프로젝트 회의진행, 문서작성, 데이터 관리 | | | |  |
| 임철성 | 샘플 모델 연구(word2vec) | | | |  |
| 손재형 | 샘플 모델 연구(word2vec) | | | |  |
| 이민흠 | 샘플 모델 연구(word2vec) | | | |  |
| 주성민 | 샘플 모델 연구(word2vec) 및 프론트 구상 | | | |  |
| 정우현 | 샘플 모델 연구(word2vec) 및 프론트 구상 | | | |  |

| **회 의 내 용** |
| --- |
| [안건 1 - 크롤링 조]  - 샘플 데이터 조 엑셀 양식에 맞게 크롤링된 결과를 전처리 필요  - 이미지를 어떤식으로 가져 왔는지? => AIHub 내에 annotation 전체를 스크린샷 했다는  얘기 => 소통 오류  - 샘플 조에서 데이터에 대한 소개 부분 어디서 가져왔는지? => AIHub에서 가져옴  - 데이터 '소개 & 구축목적'을 크롤링할 것 => 부족하면 pdf에서 추가적으로 가져올 것  - '보행 안내' 라는 문장 넣었을 때 '장애인 길 안내'가 나올 수 있도록 유사도  - 지금 할 것은 회사가 어떠한 정보를 넣었을 때 데이터와 매칭이 되기 위해 유사도 분석 필요  - 회사가 단어를 넣을 수도 있고 홈페이지 정보를 넣었을 때 매칭을 시켜서 보여줘야 함  => AI가 전혀 안들어감  - AIHub 내 380개에 데이터를 읽으면서 구성원 모두가 감이 잡혔을 것으로 예상  [안건 2 - 샘플링 조]  [세부내용]  - 우현 + 재혁(11.28.월 까지 완료)  1. 한개 분야별로 3개 데이터로 분리한 것을 웹페이지 상에서 선택해서 보여줄 수 있도록 =>  AI 불필요  2. 대/중/소분류에서 구성된 데이터 셋에서 어떤 것이 적절한지 보여줄 것 => 인터페이스에  신경쓸 것    - 모델링 부분(샘플 조 + 크롤링 조 협업 필요)  - 회사의 홈페이지 정보와 데이터를 매칭을 시킬 것인가? 단어와 단어를 매칭을 시킬 것이냐?가 문제 => 두 개 다  - 채용 공고 올라온 회사 애니모비, 한컴n플러스, +(내일 올 회사)를 포함 시켜 유사도 분석  - 코사인 유사도보단 Word2Vec을 쓰는게 좋음 => 크롤링 + 재형씨 + 성민이가 같이 할 것  [안건 3 - To-do List]  1. 타이틀, 소개, 구축목적을 크롤링해서 샘플링 조의 엑셀 파일 양식대로 맞출 것  => 철성, 민흠  2. 코사인유사도, Word2Vec, FastText 사용해서 두 개의 텍스트 사이에 점수를 내는 방법을  마련할 것(Plan A) => 재형, 성민  3. 웹페이지 관련 plan C => 우현, 재혁  4. 다중레이블 방식은 plan B이나 아이디어 부족으로 좀 더 고려  5. 각 영역별로 세부적인 계획 세울 필요 있음 |

**[세모 AI]**

|  |  | | **작성자** | 주성민 | |
| --- | --- | --- | --- | --- | --- |
|  |  | | **날 짜** | 2022. 11. 28. ( 10회) | |
| **팀 명** | 코람페(코딩해서 람보르기니, 페라리 사자) | | | | |
| **팀구성원** | **이름** | **역할** | | | |
| 이재혁 | 팀장(프로젝트 총괄 및 문서작성) | | | |
| 임철성 | 데이터 관리 및 모델링 | | | |
| 손재형 | 데이터 전처리 및 모델링 | | | |
| 이민흠 | 데이터 전처리 및 모델링 | | | |
| 주성민 | 데이터 전처리 및 모델링 | | | |
| 정우현 | 모델링, 웹 화면 구현 | | | |
| **일별 진행상황** | | | | | |
| **이름** | **업 무 내 용** | | | | **특이사항** |
| 이재혁 | 프로젝트 회의진행, 문서작성, 데이터 관리, 프론트엔드 | | | |  |
| 임철성 | 샘플 모델 연구(ko-sentence-bert) | | | |  |
| 손재형 | 샘플 모델 연구(word2vec) | | | |  |
| 이민흠 | 샘플 모델 연구(ko-sentence-bert) | | | |  |
| 주성민 | 샘플 모델 연구(word2vec) | | | |  |
| 정우현 | 샘플 모델 연구(word2vec) 및 프론트엔드 | | | |  |

| **회 의 내 용** |
| --- |
| [안건 1 - Front-end]  1. 시연 시 너무 많은 분류 목록은 심사위원을 질리게 할 수 있다.  2. HTML, CSS => 재혁 전담(디자인 부분) / JavaScript => 우현 전담(기능적인 부분)  To-do List  - 재혁 : 디자인 변경, modelling이 어디 부분에서 어떤 서비스 형태로 적용할지 강구  - 우현 : 기능적 구현(AI 모델 적용), 결과를 보여줄 수 있는 방안 강구, 디테일 살리기,  서버 만들기 및 세부내용 작성  [안건 2 - 모델링 조]  [세부내용]  1. AIHub annotation의 일부분의 정보를 어떻게 가공을 해서 보여줄지 고민 필요  => 자료를 던져주고 알아서 해석하라는 느낌을 받음  2. ko-sentence-transformer 서버 터진거 관련 유사 모델 찾아볼 필요 있음  To-do List  - 민흠  1) 참고 문헌 조사 필요  a) 우리가 쓰는 모델이 어떤 기반에서 돌아가는지 설명 필요  b) 하루를 넘기지 말 것  2) STS(문장유사도 비교 테스트셋)과 문장 유사도 관련 트랜스포머의 임베딩 레이어를  통해 알아볼 것  - 철성  1) 유사도 분석 모델 다시 테스트  2) 활용 AI 모델 및 코드에서 3개 유사데이터 추천할 때 인공지능 모델/필요성 부분 확인  3) 서버 만들기 및 세부내용 작성  4) 회사 웹페이지에서 헤더(예상하지 못한 경우 또한)를 크롤링하기  5) 회사 홈페이지 30개 정도 조사, 이를 기반으로 30개 이외에 다른 회사도 돌아갈 수  있도록  [안건 3 - 서비스 기획]  1. 100가지 많은 정보를 주면 소비자가 안씀. => 최적의 정보를 줄 수 있도록  2. 회사의 홈페이지 정보 분석을 끌어내서 워드 임베딩을 통해 가장 잘 맞는 데이터를  추천하는 방향  => 회사의 홈페이지만 넣어두면 우리가 알아서 적절한 매칭을 제공하는 서비스  3. 홈페이지마다 헤더 정보를 통해 우리가 찾아야하는 정보가 무엇인지 파악을 해야함  => 텍스트로 끌고 올 수 있는 회사 위주로  4. Pro에서는 정확도의 근거(더 많은 정보, 맞춤형 결과를 내놓는 것)을 위주로 구성해야함  => 즉, 자동화 & 맞춤형 |

**[세모 AI]**

|  |  | | **작성자** | 주성민 | |
| --- | --- | --- | --- | --- | --- |
|  |  | | **날 짜** | 2022. 11. 30. ( 11회) | |
| **팀 명** | 코람페(코딩해서 람보르기니, 페라리 사자) | | | | |
| **팀구성원** | **이름** | **역할** | | | |
| 이재혁 | 팀장(프로젝트 총괄 및 문서작성) | | | |
| 임철성 | 데이터 관리 및 모델링 | | | |
| 손재형 | 데이터 전처리 및 모델링 | | | |
| 이민흠 | 데이터 전처리 및 모델링 | | | |
| 주성민 | 데이터 전처리 및 모델링 | | | |
| 정우현 | 모델링, 웹 화면 구현 | | | |
| **일별 진행상황** | | | | | |
| **이름** | **업 무 내 용** | | | | **특이사항** |
| 이재혁 | 프로젝트 회의진행, 문서작성, 데이터 관리, 프론트엔드 | | | |  |
| 임철성 | 샘플 모델 연구(ko-sentence-bert) | | | |  |
| 손재형 | 샘플 모델 연구(word2vec) | | | |  |
| 이민흠 | 샘플 모델 연구(ko-sentence-bert) | | | |  |
| 주성민 | 샘플 모델 연구(word2vec) | | | |  |
| 정우현 | 샘플 모델 연구(word2vec) 및 프론트엔드 | | | |  |

| **회 의 내 용** |
| --- |
| [안건 1 - Front-end]  - 키워드가 맞춤형이 들어가야 함  - 맞춤형 + AI 중심으로 Depth을 추구할 것  - 4가지(간편하게, 맞춤형, AI, 추가 정보)가 어우러져서 들어갈 수 있도록 할 것  [안건 2 - 모델링 조]  [세부내용]  - 1) Ko-Sentence-Transformer 사용 모델링 = 구동확인(ko-sroberta-multitask)  - 2) STS와 문장 유사도 관련 트랜스포머의 임베딩 레이어를 통해 알아보기(확인중)  - 3) 활용 모델 및 코드에서 3개 유사데이터 추천할 때 인공지능 모델/필요성 부분 확인  (현재는 추가 해놓은 상태 = 추후 재판단 필요(AI모델링 패키지 선택인원들에게 기본정보로 제공하는 것도 나쁘지 않을듯))  \* AI모델 개발간 참고를 위한 자료정도로 AI모델 개발 회사에게 제출하는 형태  [안건 3 - 모델링 2]  - 평균을 내면 값의 차이가 거의 안날 것임  - 또 다른 아이디어를 낼 필요 있음  - 경영정보학 논문을 끊임없이 읽어봐야 함  (학회지 - 어느 학회지에 주로 내는지 확인 : 약 100개)  => 졸업 했을 때 기준 아이디어가 딱 떠오를 수 있을 정도로... |

**[세모 AI]**

|  |  | | **작성자** | 주성민 | |
| --- | --- | --- | --- | --- | --- |
|  |  | | **날 짜** | 2022. 12. 02. ( 12회) | |
| **팀 명** | 코람페(코딩해서 람보르기니, 페라리 사자) | | | | |
| **팀구성원** | **이름** | **역할** | | | |
| 이재혁 | 팀장(프로젝트 총괄 및 문서작성) | | | |
| 임철성 | 데이터 관리 및 모델링 | | | |
| 손재형 | 데이터 전처리 및 모델링 | | | |
| 이민흠 | 데이터 전처리 및 모델링 | | | |
| 주성민 | 데이터 전처리 및 모델링 | | | |
| 정우현 | 모델링, 웹 화면 구현 | | | |
| **일별 진행상황** | | | | | |
| **이름** | **업 무 내 용** | | | | **특이사항** |
| 이재혁 | 프로젝트 회의진행, 문서작성, 데이터 관리, 프론트엔드 | | | |  |
| 임철성 | 샘플 모델 연구(ko-sentence-bert) | | | |  |
| 손재형 | 샘플 모델 연구(word2vec) | | | |  |
| 이민흠 | 샘플 모델 연구(ko-sentence-bert) | | | |  |
| 주성민 | 샘플 모델 연구(word2vec) | | | |  |
| 정우현 | 샘플 모델 연구(word2vec) 및 프론트엔드 | | | |  |

| **회 의 내 용** |
| --- |
| [안건 1 - Front-end]  - 유료화의 장점에 대한 서술  - Basic과 Pro 사이의 특징이 구체적으로 나오게 할 것  - 결제창은 이미지만 띄울 것  - 모델링 조에게 아웃풋이 어떤 식으로 나올 지 잘 배치를 해야함  [안건 2 - 모델링 조]  - 회사 홈페이지 헤더를 크롤링하기  (실패 = 회사마다 구성이 전부 다르며 일부회사들의 경우에는 메뉴위에 마우스를  올려야지만 메뉴가 보이는등 제한사항이 많음)  - 회사 홈페이지 30개정도 조사 후 이를 기반으로 30개 이외의 다른 회사도 들어갈 수  있도록(실패)  - 4~5번 실패에 따른 대체 방식으로 사용자에게 회사명을 입력받을시 '커리어'사이트에서  검색 후 주요사업내용을 크롤링해와 transformer모델을 통해 유사도 분석 실시후 적합한  데이터셋 추천    - 검색이 제한되는 회사의 경우 사용자에게 직접 주요사업에 대하여 추가입력을 요구하는 형태로 구현  [안건 3 - 모델링 조 2]  - 평균내는거 말고 어떻게 처리할 것인가가 관건임  (기준 : 길이가 긴 경우 - 차별화 될 수 있는 스코어를 뽑아 내는거)  - 무엇이 종속변수가 될 지 파악을 하고 독립변수가 무엇이 될 지의 직관을 깨닫길  - 수학적 베이스 필요(어려운 수학이 아닌 간단한 수학)  - 동시에 Word Embedding 모델 만들기  - 알고리즘 짜고(재형) / 맞춤형 임베딩 모델(성민) 만들기 |

**[세모 AI]**

|  |  | | **작성자** | 주성민 | |
| --- | --- | --- | --- | --- | --- |
|  |  | | **날 짜** | 2022. 12. 05. ( 13회) | |
| **팀 명** | 코람페(코딩해서 람보르기니, 페라리 사자) | | | | |
| **팀구성원** | **이름** | **역할** | | | |
| 이재혁 | 팀장(프로젝트 총괄 및 문서작성) | | | |
| 임철성 | 데이터 관리 및 모델링 | | | |
| 손재형 | 데이터 전처리 및 모델링 | | | |
| 이민흠 | 데이터 전처리 및 모델링 | | | |
| 주성민 | 데이터 전처리 및 모델링 | | | |
| 정우현 | 모델링, 웹 화면 구현 | | | |
| **일별 진행상황** | | | | | |
| **이름** | **업 무 내 용** | | | | **특이사항** |
| 이재혁 | 프로젝트 회의진행, 문서작성, 데이터 관리, 프론트엔드 | | | |  |
| 임철성 | 샘플 모델 연구(ko-sentence-bert) | | | |  |
| 손재형 | 샘플 모델 연구(word2vec) | | | |  |
| 이민흠 | 샘플 모델 연구(ko-sentence-bert) | | | |  |
| 주성민 | 샘플 모델 연구(word2vec) | | | |  |
| 정우현 | 샘플 모델 연구(word2vec) 및 프론트엔드 | | | |  |

| **회 의 내 용** |
| --- |
| [안건 1 - Front-end]  - 플라스크 서버 구축 완료  - 플라스크와 모델 연동 필요 => (강사님) os 명령어를 통해서 python을 기동  [안건 2 - 모델링 조]  - 평균값 모델 구축 완료  - 동일 사례 입력을 했을 때에도 유사도 값이 적게 나오는 문제 식별  - PPT 도식화 필요 => 발표 시에도 간략하게 도식화 후 설명 필요  - 단어가 여러개 있어서 평균을 내면 똑같아 짐 => 가장 높은 값 3개만 추려내서 합을 내거나 평균값을 낼 필요 있음  - AIHub 내 추가 학습 시 형태소 분석 필요 => 500메가 정도가 2시간 정도 걸림 / 1기가 4시간  - 미캡이 형태소 분석 성능이 좋으면서 빠르다. 그러나 서술어를 잘 분석을 못한다.  [안건 3 - 모델링 조 2]  - 함수 형태로 제작 완료  - 웹 구현 테스트 중 |

**[세모 AI]**

|  |  | | **작성자** | 주성민 | |
| --- | --- | --- | --- | --- | --- |
|  |  | | **날 짜** | 2022. 12. 07. ( 14회) | |
| **팀 명** | 코람페(코딩해서 람보르기니, 페라리 사자) | | | | |
| **팀구성원** | **이름** | **역할** | | | |
| 이재혁 | 팀장(프로젝트 총괄 및 문서작성) | | | |
| 임철성 | 데이터 관리 및 모델링 | | | |
| 손재형 | 데이터 전처리 및 모델링 | | | |
| 이민흠 | 데이터 전처리 및 모델링 | | | |
| 주성민 | 데이터 전처리 및 모델링 | | | |
| 정우현 | 모델링, 웹 화면 구현 | | | |
| **일별 진행상황** | | | | | |
| **이름** | **업 무 내 용** | | | | **특이사항** |
| 이재혁 | 프로젝트 회의진행, 문서작성, 데이터 관리, 프론트엔드 | | | |  |
| 임철성 | 샘플 모델 연구(ko-sentence-bert) | | | |  |
| 손재형 | 샘플 모델 연구(word2vec) | | | |  |
| 이민흠 | 샘플 모델 연구(ko-sentence-bert) | | | |  |
| 주성민 | 샘플 모델 연구(word2vec) | | | |  |
| 정우현 | 샘플 모델 연구(word2vec) 및 프론트엔드 | | | |  |

| **회 의 내 용** |
| --- |
| [안건 1 - Front-end]  - 서비스 미리보기 부분 논의 중 => 마무리 필요  - 멘토님이 디자인 부분 보완  [안건 2 - 모델링 조]  - 입력문의 형태소 분석은 Rhino로 기존대로 진행  - 솔트룩스는 자체 형태소 분석기를 갖고 있기 때문에 돌리지 못한다...  - 1기가 정도로 학습 시킬 수 있게끔...  [안건 3 - 모델링 조 2]  - 플라스크 연동 필요 => 결과가 나오는 것 확인 |

**[세모 AI]**

|  |  | | **작성자** | 주성민 | |
| --- | --- | --- | --- | --- | --- |
|  |  | | **날 짜** | 2022. 12. 08. ( 13회) | |
| **팀 명** | 코람페(코딩해서 람보르기니, 페라리 사자) | | | | |
| **팀구성원** | **이름** | **역할** | | | |
| 이재혁 | 팀장(프로젝트 총괄 및 문서작성) | | | |
| 임철성 | 데이터 관리 및 모델링 | | | |
| 손재형 | 데이터 전처리 및 모델링 | | | |
| 이민흠 | 데이터 전처리 및 모델링 | | | |
| 주성민 | 데이터 전처리 및 모델링 | | | |
| 정우현 | 모델링, 웹 화면 구현 | | | |
| **일별 진행상황** | | | | | |
| **이름** | **업 무 내 용** | | | | **특이사항** |
| 이재혁 | 프로젝트 회의진행, 문서작성, 데이터 관리, 프론트엔드 | | | |  |
| 임철성 | 샘플 모델 연구(ko-sentence-bert) | | | |  |
| 손재형 | 샘플 모델 연구(word2vec) | | | |  |
| 이민흠 | 샘플 모델 연구(ko-sentence-bert) | | | |  |
| 주성민 | 샘플 모델 연구(word2vec) | | | |  |
| 정우현 | 샘플 모델 연구(word2vec) 및 프론트엔드 | | | |  |

| **회 의 내 용** |
| --- |
| [안건 1 - Front-end]  - 그 외 나머지 부분 구현 완료  - Basic과 Pro의 차이? 질문이 나올 것으로 예상 => 효과적으로 충분히 설명될 수 있도록 홈페이지 구현  [안건 2 - 모델링 조]  - sentence 부분만 끌고와서 하나의 텍스트 파일로 만들어야 함 => wiki 데이터, nsmc 데이터 사용해서 먼저 확인 후 => sentence 부분 추출하기  - json 파일 접근할 때는 json 파일을 접근하는 도구가 있는데 sentence를 뽑아올 수 있도록 => json은 api가 따로 있음 => json api 이용 필요  - 진행하면서 강사님과 상의 필요  [안건 3 - 모델링 조 2]  - 엉성하더라도 하나의 웹페이지에 구현이 될 수 있도록 => 이후 정교화 작업 |

**[세모 AI]**

|  |  | | **작성자** | 주성민 | |
| --- | --- | --- | --- | --- | --- |
|  |  | | **날 짜** | 2022. 12. 12. ( 14회) | |
| **팀 명** | 코람페(코딩해서 람보르기니, 페라리 사자) | | | | |
| **팀구성원** | **이름** | **역할** | | | |
| 이재혁 | 팀장(프로젝트 총괄 및 문서작성) | | | |
| 임철성 | 데이터 관리 및 모델링 | | | |
| 손재형 | 데이터 전처리 및 모델링 | | | |
| 이민흠 | 데이터 전처리 및 모델링 | | | |
| 주성민 | 데이터 전처리 및 모델링 | | | |
| 정우현 | 모델링, 웹 화면 구현 | | | |
| **일별 진행상황** | | | | | |
| **이름** | **업 무 내 용** | | | | **특이사항** |
| 이재혁 | 프로젝트 회의진행, 문서작성, 데이터 관리, 프론트엔드 | | | |  |
| 임철성 | 샘플 모델 연구(ko-sentence-bert) | | | |  |
| 손재형 | 샘플 모델 연구(word2vec) | | | |  |
| 이민흠 | 샘플 모델 연구(ko-sentence-bert) | | | |  |
| 주성민 | 샘플 모델 연구(word2vec) | | | |  |
| 정우현 | 샘플 모델 연구(word2vec) 및 프론트엔드 | | | |  |

| **회 의 내 용** |
| --- |
| [안건 1 - Front-end]  - 웹 영상 촬영 수정(1분)  - 웹 페이지 수정정  [안건 2 - 모델링 조]  - Word2Vec 관련 PPT 내용 구성  - 발표 대본 수정  [안건 3 - 모델링 조 2]  - S-BERT 특/장점 분석 및 PPT 반영영 |