## **데이터 수집** SKN Family Al Camp 2기 : 최종 프로젝트 2조



| 프로젝트 주제      | 경제 지식이 없는 아동 및 국민들을 대상으로 경제 지식을 쉽게 설명하는 AI 인플루언서 웹 페이지 관리 Tool  |   |   |                        |  |
|--------------|---|---|---|------------------------|--|
|              |   |   |   | 데이터                    | 경제금융용어 700선(375p),   |
| 데이터 출처       | 한국은행  | 데이터 형태  | pdf 파일  | 정보 및 분량<br>(사이즈)       | 알기 쉬운 경제이야기 : 일반인용 (335p)  |
|              | 아동 및 청소년을 대상으로 하는 정보전달 목적의 AI 인플루언서 구축이 목표이기 때문에 정확한 정보를 학습시키는 것이 중요하다고 생각했습니다. 한국은행은 자본이 개입되지 않은 독립적인 특수 법인이라 편향성이 거의 없는 데이터를 제공해 줄 수 있을 것으로 기대했습니다. 한국은행은 일반 대중들을 위해 전자책 형태의 pdf 문서를 제공합니다. 저희 팀은 그 중에서 '경제금융용어 700선'과 '알기 쉬운 경제 이야기(일반인용)'을 선택했습니다.  알기 쉬운 경제 이야기 시리즈 중에서 초/중/고등학생 용도로 쓰여진 것을 선택하지 않은 이유는 정보 압축률 때문입니다. 초/중/고등학생 입장에서 서술한 자료는 쉽게 설명하기 위해, 단순 개념을 타겟으로 길게 서술합니다. 저희는 AI에게 압축적으로 지식과 맥락을 학습시키기 위 |   |   |                        |  |
|              |   |   |   |                        |  |
| 데이터<br>수집 계기 |   |   |   |                        |  |
|              | AI에 최대한 압축적으로 경제 지식을 습득시킨 후에 파인튜닝과 프롬프트 엔지니어링을 활용하여 읽기 쉬운 서술체로 변환시킴으로써, 고난<br>도의 지식을 알고 있지만 쉽게 설명할 수 있는 '경제 정보 제공 목적의 AI 인플루언서'를 제작할 수 있을 것으로 기대합니다.  |   |   |                        |  |
|              | 이와 같은 판단 하에 '경제금융용어 700선'과 '알기 쉬운 경제이야기 : 일반인용' 전자책 pdf 문서를 사용 데이터로 선정했습니다.   |   |   |                        |  |
| 데이터<br>출처    | [경제금융용어 700선] https://www.bok.or.kr/portal/bbs/B0000216/view.do?nttld=10075809&type=&searchOptn8=22&menuNo=200134&listType=G&pageIndex=1  [알기 쉬운 경제이야기: 일반인용] https://www.bok.or.kr/portal/bbs/B0000249/view.do? nttld=10061816&searchCnd=1&searchKwd=&depth2=201157&depth3=201269&depth4=200765&depth=200765&pageUnit=10&pageIndex=1&programType=news Data&menuNo=200765&oldMenuNo=201269                        |   |   |                        |  |
|              |   |   |   |                        |  |
| 원본<br>데이터    | 대신 계파간 다시<br>결과 비즈가의 등<br>의 인도 당하는<br>강안하면 사실상<br>당신으로 운전의<br>당신으로 운전을<br>자신에 사는 사용기에<br>가는 사용함이 때<br>경우 운영의 때<br>경우 운영의 때<br>경우 운영의 이 임금임성이 :<br>교객확인절차<br>(어디) 양공원 기위<br>이 일본 기위에<br>(시조 소리에서<br>기계 기계 기   |   | 관 등 증권거리의  호 전한 설품관련  소개 선통 등을 기관과 계곡대체 원이 제공간 대체 모드는 지난 경기 10년 개의 10년 기계 전한 10년 기계 지난 10년 기계 전한 10년 기계 전 |                        | 가합니다. 이것은 꼭 정치인이나 공리자들이 압반인보다 원만 더 뛰어난 등대를 가졌다는 별은 아니지만 사람들이 선택한 정부의 약층을 받고 나라 경제 운영을 받습니다. 그래서 나라장계획을 참 운영할 수 있는 충분한 실차인가 공의자를 신흥하는 것이 중요하며 이를 위해서는 경계에 대한 음박은 이해대를 갖춘 사면이 요구되니다. 한 나라의 반두수의 수준이 전면의 회의에서 수준에 달려 있는 경제에 대한 경반을 수준에 비해한 나고 볼 수 있습니다. 국민경제와 정부생태를 제대로 이해하고 생가할 수 있는 경제에 대한 경소를 설치는 전기를 제공하게 하면 경기를 보고 있습니다. 국민경제와 정부생태를 제대로 이해하고 생가할 수 있는 경인에 대한을 갖는는 것은 나라장계의 발산에 의원하고 발가할 수 있는 경안나다. 국민의로가운 관심을 받는 원드컵 속구경기에서 중은 심하를 내려면 경기를 위한 생각이 되었다. 경기를 가는 것은 발생이 발산에 의원하고 발생이 되었다. 전체 기본 등 신화에 설치를 받게 함께 되었다. 이를 의용하는 등은 인교 속구를 들려면 생대에 의원한 것이 함께 대한 경기에 의원하는 방법이 하나가 보고 모음 등에 사람들이 안나가 내고 발생이 되었다. 이를 의성하는 방법이 하나가 보고 모음 등에 사람들이 안나가 내고가 바라고 발생이 학생을 받는 기는 것이나다. 경제교육을 통해 사람들이 안나가 내고가 바라고 발생이 학생을 받아 보았다라면 궁극하고로 국민생계가 함께 되어 의원생을 받아 회생되는 말로 발생하게 함께 적으로 운영하고 보다 바라의 함 심과를 얻는 보내가 바란된 것입니다.  2. 생활은 선택의 연속 - 선택은 행복의 최대화 과정 |
|              | 인구(노동가능인  | 창에서 매월 작성하고 있는 경제활동인구조사에서 :<br>귀에 대해 취업자가 차지하는 비율을 말한다. 건권 설<br>J자가 차지하는 비율을 말한다. 고용론을 설업률의 | 일업률은 경계활동   | 배하고, 소비하는 모든<br>명명을 받던 | 입상생활은 청계활동의 연속입니다. 일반적으로 못과 음식, 방원고의 자<br>등적 같은 재위나 관광, 금융, 오락과 같은 서비스를 생산, 분백, 소비하는<br>것을 모두 경제활동이라고 한 수 있습니다. 살아가는 동안에 가츠적인 생<br>활을 영위하고 또 그대그때 여러 가지 유방을 충족시키기 위해서는 여러  |

## 데이터 전처리 시 유의해야 할 점

'알기 쉬운 경제이야기' 문서의 경우, 날개 부분에 본문 내용을 보충하는 용어 설명 파트가 있습니다. 그대로 pdf 문서를 읽어 올 경우에 본문 내용과 날개 속 용어 설명이 섞일 가능성이 존재합니다. 따라서 각 부분을 분리해 주는 작업을 수행해야 합니다.

< 알기 쉬운 경제이야기: 일반인용 >

< 경제금융용어 700선 >

'경제금융용어 700선' 문서는 단순 사전 형태로 구성되어 있어서 전처리할 부분이 비교적 적은 편에 속하지만, '알기 쉬운 경제이야기'는 여백에 용어 설명을 서술하고 있으므로 전처리 시 고려해야 할 사항이 확고하게 존재합니다.