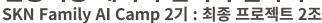
인공지능 데이터 전처리 결과서





프로젝트 주제

경제 지식이 없는 국민들을 대상으로 경제 지식을 쉽게 설명하는 AI 인플루언서 웹 페이지 관리 Tool

사용 데이터

[한국은행 제공 경제교육용 PDF 문서]

- _ 경제금융용어 700선
- _ 알기 쉬운 경제 이야기 (일반인용)

[원본]



'경제금융용어 700선'은 크게 용어 명칭, 용어 설명, 연관 검색어로 구성되어 있습니다.

'경제금융용어 700선'

전처리 과정 (1)

[사용 Library & Tool]



PDF 속 텍스트 추출 시에 langchain_unstructured 라이브러리를 사용했습니다. 해당 라이브러리는 pdf와 word 문서에서 text를 추출하는 데에 강점을 갖고 있습니다. 또한 metadata 속성을 통해 문서 구조에 접근할 수 있다는 장점이 존재합니다.

| pandas

추출한 텍스트는 판다스 라이브러리를 사용하여 CSV 문서 형태로 정리했습니다. 용어 명칭, 용어 설명, 연관 검색어를 필드 형태로 가공한 후에 데이터프레임 형태로 정리했습니다.

[1차 전처리 : PDF 속 텍스트 추출 (langchain_unstructured 라이브러리)]

- 1) PDF 파일 내용을 텍스트로 변환
- 2) metadata 속성이 'Title'이거나 'NarrativeText'인 항목만 추출
- 3) 목차에 해당하는 자음 요소를 제거

[1차 전처리 결과물 : 텍스트 파일]

가계수지

가정에서 일정 기간의 수입(명목소득)과 지출을 비교해서 남았는지 모자랐는지를 표시한 것을 가계수지(household's total income and expenditure)라 한다. 가계수지가

혹자를 냈다면 그 가정은 벌어들인 수입 일부만을 사용했다는 것을 의미하며, 적자를 냈다면 수입 외에 빚을 추가로 얻어 사용한 것이라고 보아야 한다. 우리나라는 통계청에 서 가계의 수입과 지출을 조사하여 국민의 소득수준 및 생활실태를 파악하기 위해 표본으로 선정된 가계에 가계부를 나누어 주고 한 달간의 소득과 지줄을 기록하도록 한 다음 이를 토대로 가계수지

기계구의 구하기구요 단 글인크 국도의 지글은 기록이로록 단 다음 하르 포테포 기계구시 통계를 작성하여 발표하고 있다. 가계부의 소득항목에는 근로소득 사업소득 재산소득 이전소득 항목이 있고, 비용항목에는 식료품비 주거비 수도광열비 보건의료비 교육비 항목이 있다. 연관검색어 : 경상수지, 재정수지 [2차 전처리 : CSV 문서 형태 가공 (Pandas 라이브러리)]

- 1) 용어 명칭(title), 용어 설명(content), 연관 검색어(related_keyword)로 분리하여 딕셔너리로 저장
- 2) 딕셔너리를 데이터프레임으로 변환
- 3) 데이터프레임을 CSV 문서 형태로 저장

[2차 전처리 결과물: CSV 문서]

'경제금융용어 700선'

전처리 과정 (2)

```
title content
기계부심위함자* 가구의 소득 조름은 물론 금융 및 실물 자산까지 중합적으로 고려하여 가계부제의 부실위험을 참가하는 지표로, 가게의 채무산환능력을 소득 증부계환리공상환비율(DSR)
기계수지
기계수지점을 보면적으로 자주를 주문수 경력을 제본하는소등으로 나는 배트를 받다나는 통계이다. 가게신자의 가게수지(Mouseholds total income and expenc 경상수지, 제공수지
기계수지점을 보면적으로 자주를 주문수 제품을 제본하는소등으로 나는 배트를 얻었다는 미모자장는지를 표시한 것을 가게수지(Mouseholds total income and expenc 경상수지, 제공수지
기계수지점을 보면적으로 자주를 주문수 제품을 제본하는소등으로 나는 배트를 얻었다는 미모자장는지를 표시한 것을 가게수지(Mouseholds total income and expenc 경상수지, 제공수지
기계수지점 등 기계신으를 가게신으를 가게 기계부로 대한 1년 등을 급규모로 나타나는 통계이다. 가게신용은 급용기관보로 아니라 경보, 반면되자 등 기타기관이 기계에 제공한 대출을 의상구의 관련 신용을 모두 포함하는 것으로 크게 가게 다음과 만든 기계점 제공한 대출을 기실수로 되었는지 말로 발생으로 보면 기계적으로 기계적으로 보면 기계적으로 기계적으로 보면 기계적으로 보면 기계적으로 기계적으로 보면 기계적으로 보는 고객을 진사이다면 구리적으로 이연 기계적으로 보면 기계적으로 보는 고객을 진사이라고 보급이 기계적으로 보면 기계적으로 기계적으로 보다 고객을 보면 기계적으로 보다 기계적으로 보다 고객을 보면 기계적으로 보는 고객을 지수 기계적으로 보는 고객을 지수 기계적으로 보는 고객을 지수 기계적으로 보는 고객을 지수 기계적으로 보다 고객을 보면 기계적으로 보다 고객을 기계적으로 보는 고객을 지수 기계적으로 보다 고객을 보면 기계적으로 보다 고객을 보면 기계적으로 보다 고객을 보면 기계적으로 보다 고객을 기계적으로 기계적으로 보다 기계적으로 보다 기계적으로 보다 고객을 기계적으로 보다 기계적으로 보다 보면 기계적으로 보다 기계적으로 보다 기계적으로 보다
```

2차 전처리까지 마친 CSV 문서의 데이터 사이즈는 700 rows x 3 columns 입니다.

[원본]



'알기 쉬운 경제 이야기 (일반인용)'

전처리 과정 (1)

'알기 쉬운 경제 이야기(일반인용)'은 크게 제목, 본문, 그림, 날개 용어 설명으로 구성되어 있습니다.

[사용 Library & Tool]



날개 용어 설명을 제거하기 위해서 pilow 라이브러리를 사용했습니다. 짝수 페이지와 홀수 페이지 각각의 특징을 반영하여 날개 용어 설명을 제거한 후, 이미지로 저장했습니다.



날개 용어 설명을 지운 이미지에 CLOVA OCR을 적용함으로써 내용을 추출했습니다. 추출한 내용은 텍스트 파일 형태로 저장했습니다.

| pandas

추출한 텍스트 내용은 판다스 라이브러리를 사용하여 CSV 문서 형태로 정리했습니다. 제목(title), 본문(content)를 필드 형태로 가공한 후에 데이터프레임 형태로 정리했습니다.

[1차 전처리 : PDF 속 텍스트 추출 (PIL 라이브러리, CLOVA OCR)]

- 1) 날개 용어 설명 제거된 형태로 이미지 저장
- 2) OCR 적용

[1차 전처리 결과물 : 텍스트 파일]

생활과 경제

1. 왜 경제이해력이 중요한가?

우리 삶의 많은 부분은 경제와 연관

대부분의 사람들은 행복하게 살기를 원합니다. 행복한 삶을 이루기 위해 서는 건강한 가운데 가정과 직장 그리고 사회생활에서 적절한 성취를 얻는 것이 무엇보다 중요합니다. 여기에 검제적으로 문택한 것도 행 목한 삶의 여러 가지 조건 중 하니라고 할 수 있습니다. 현대의 물질문명에 대해 비판 이 없는 것은 아 니지만 검제적 품요가 협복의 한 요소인 것만은 틀림없습니다.

행복을 추구하는 우리 삶의 많은 부분은 경제적 의사결정과 밀접하게 관련되어 있습니다. 우리는 일상생 활에서 의사결정을 계속하지 않을 수 없습니다. 예를 들어 시장에서 장을 볼 때 일반 체소를 살 것인지 조금 비싸 더라도 유기농 체소를 살 것인지를 비롯해 막시를 알까 버스를 할까와 갈 이 대중교통을 이용 할 때의 사소한 결정에서부터 대화직한 여부, 직접 선택, 주택 구입 등 리고 중요한 결정에 이르기까지 어떤 결정을 해야만 할 때가 해야될 수 없이 많습니다. 이러한 의사결정을 할 때 경제상식과 지식 을 적 결히 활동한 경제적 사고방식은 현명한 의사결정을 내리는 데큰 도움이 됩니다.

그렇다고 경제지식이 사람들을 골바로 부유하게 해주는 것은 아닙니다. 경제학 강의 첫 시간에 교수들이 강조하는 말은 "경제학이 부자가 될수 있게 돈 버는 방법을 가르치는 학문은 아니다. 라는 것입니다. 그 렇다면 경제지식이 왜 필요할까요? 과연 이 책을 읽기 위해 투자한 시간의 가치에 상당하는 결설이 돌아 물까요? 경제성식이나 지식이 많다고 해서 모든 사 람이 곧 부자가 되고 행복하질 수 있다면 참 좋을 것 입니다. 그러나 현실은 그렇지 않습니다. 역사적으로 저명한 경제학자라고 해서 반드시 부자가 아 니었던 것은 이것을 잘 대변해줍니다.

경제상식이나 지식이 우리를 한 순간에 부자로 만들어 주지는 않습니다. 그렇지만 우리가 일상사에서 합 리적인 의사결정을 할수 있도록 도와줌으 로써 장기적으로 보면 개개인의 행복을 크게 해주는 역할을 합 니다. 주식 투자를 예로 들어 보겠습니다. 경제지어이 단기적으로 주식으로 돈을 버는 기법을 가르쳐 주 지는 않습니다. 그렇지만 위원을 분산하고 경제의 흐름을 이해할 수 있도록 함으로써 충짐기적으로 순실 을 최소화하고 나이가 재산을 늘릴 수 있도록 하는 데 도움을 줄 수 있습니다.

[2차 전처리 : CSV 문서 형태 가공 (Pandas 라이브러리)]

- 1) 제목(title), 용어 설명(content)으로 분리하여 딕셔너리로 저장
- 2) 딕셔너리를 데이터프레임으로 변환
- 3) 데이터프레임을 CSV 문서 형태로 저장

[2차 전처리 결과물 : CSV 문서]

'알기 쉬운 경제 이야기 (일반인용)'

전처리 과정 (2)

title	content
우리 삶의 많은 부분은 경제와 연관	대부분의 사람들은 정부하게 생기를 원합니다. 정부한 삶을 이루기 위해 서는 건강한 가운데 가장과 직장 그리고 사실성함에서 직접한 성취를 얻는 것이 다 행복을 추구하는 우리 살인 당은 부분은 경제적 이사관점과 열장하게 관련되어 있습니다. 로리는 일상상황에서 이사관점을 보다지 않을 수 없습니다. 그렇다고 경제지식이 사람들을 판비로 부두하게 해주는 것은 아닙니다. 경제학 것의 첫 시간에 교수들이 갖춘하는 말은 경제학이 부자가 될 수 있게 준비 경제상식이나 지식이 우리를 함 순간에 부가로 만들어 주지는 않습니다. 그렇지만 우리가 열산시에서 멀리적인 의사관점을 받았다. 목모드 당근점을 받는 것이 되었다.
국민경제 운영 성과는 국민의 경제이해력에 비례	이러한 경제성식과 지식이 세계경제와 나타경제에서 여러 가지 경제한 상의 호통과 설템, 그 인과관계 등을 이해하는데 매운 도등이 되는 것은 도말함 나 일반적으로, 사람들은 개인도막 정부가 더 당기적인 시작을 갖고 나라를 운영한다고 상략합니다. 이것은 꼭 정치인이나, 공작자의 당반없다도 발표 그래서 나라정체를 잘 운영할 수 있는 종류한 정치인과 공작자를 선물하는 것이 중요하며 이용 위해서는 경제에 대한 공반은 이해력을 갖춘 시민이 요구당 국민의 뜨거운 진상을 받는 용도한 국구강기에서 중은 성과를 내려면 경기를 뛰는 선수들의 역량은 물론이고 국구를 돌관하게 외역당 짓 탓만했다.
경제활동은 선택의 연속 - 선택은 행복의 최대화 과정	임상상품은 경제활동의 연속입니다. 일반적으로 웃고 음식, 남장고와 자동자 같은 재화나 관광, 금융, 오락과 같은 서비스를 생산, 분배, 소비하는 것을 모 정제활동은 집회없는 선택을 모구합니다, 개인의 관점에서 보면 행복은 스비의 양과 전에 비리한다고 불수 있습니다. 그리나를 통해 자신의 유전에서 보면 전략으로 사용하는 것을 수 있습니다. 그리나를 통해 자신의 유전하고 공자로 인간의 육망을 채울 수 있는 자원이 복투한 것을 경제동어로 '최소성'이라고 표현되니다. 예전에는 공기나 물을 최소하지 않은 것으로 성격하고 공자로 이자 강제문제는 이렇게 이용 가능한 자원이 최소하기 때문에 성입니다. 즉 사람들이 각자 한하는 것을 다 가질 수 없다는 데에서 경제문제는 출발합니다. 다음
선택과 포기는 언제나 함께 발생	우리는 경제생활에서 회소성 때문에 선택을 하며 살고 있는데, 모든 선택에는 항상 무엇인가를 포기해야 한다는 대가가 뒤따릅니다. 즉 선택과 선택의 대기에를 들어 일을 하려면 노는 것을 포기해야 합니다. 주부가 생활비를 쓸 때도 마찬가지입니다. 돈을 가족의 건강을 위하여 좋은 음식을 만드는 데 사용할 수
하나를 가지면 다른 하나를 잃는 기회비용	모든 경제적 결정에는 대기를 지급해야 합니다. 따라서 최선의 선택을 하려면 각각의 다른 선택 대안에 대하여 이득과 손실, 편약과 비용을 따져 보고 서로 "세성에 공과 정실은 없다" 또는 '선물까 업공라다 집로과 참고다'와 같은 말이 있는데 이는 기회에 당의 의미를 잘 말해 줍니다. 우리는 선택을 받っ 따면 선 자축도 하나의 경제적인 선택입니다. 어려울 때를 대비하거나 목돈을 만들기 위한 저죽은 감과적으로 미래 소비생활에 쓰여게 됩니다. 따라서 저축은 미래
한계적으로 생각하여야 합리적인 의사결정	앞에서 우리는 회소성 때문에 발생하는 경제문제를 해소하기 위하여 편약과 비용을 비교해 합리적인 선택을 해야 한다고 하였습니다. 여기에서는 이에 덧 다른 예를 들어 보였습니다. 어느 형용시기 예약 없이 광형에서 대기하다가 비량기를 타려는 사람에게 항공료를 살펴 받아야 하는지 결혼하고 짧ん다. 3 위의 사례에서와 같이 선택에 따르는 추가적인 행복과 추가적인 회상을 비교하여 선택하는 것이 경제적 사고방식에 의한 의사경정인 것입니다. 설제 생활
유인이 의사결정에 미치는 영향은?	한편 경제주체들이 편익과 비용 개념을 한계적으로 고려하여 의사결정을 할 때, 주변 여건이 바뀌면 의사결정에 영향을 받습니다. 이처럼 의사결정에 영향 나라경제에 있어서도 마찬가지입니다, 원이나는 복게되 지율로 인해 개정적자가 누적되어 정부가 재정 군행을 예약되어 되었다. 원병을 인상하는 반 마라서 선택하는 대중요한 고려요소가 되는 건축 구체, 조세제도 동의 유인을 변화시킬 때에는 이 변화에 대한 경제주체의 방향을 미리 중도에 두당이 두당야 합
수요와 공급은 모든 경제현상을 이해하는 필수 개념	수요는 소비자가 어떤 상품을 일정한 가격을 치르고 사려고 하는 의도를 말합니다. 공급은 생산자가 어떤 가격에 판매하고자 하는 의도를 말합니다. 수요국 마찬가지로 공급은 이용을 최대화하려는 생산자의 합리적 의사점을 결과이는 어둠 가격과 생산내용 등의 움직임에 따라 변합니다. 일반적으로 각기 다른 이와 같이 수요와 공급은 경제왕성을 이해하는 데 필요한 야구 기본적인 개념인 동시에 꼭 알아야 하는 중요한 본석수단인 다음
미시경제는 나무, 거시경제는 숲	경제문제는 크게 미시경제와 거시경제의 두 측면에서 분석할 수 있습니다. 가계나 기업의 의사결정 등과 같이 개별 경제주체의 관점에서 경제행위를 분석 어떤 산의 특징을 파악하기 위해서는 숲과 나무의 모습을 모두 알아야 하듯이 이두 분야를 모두 잘 이해할 필요가 있습니다. 마잔가지로 나라경 제도 서로
교환의 이역과 가격의 신비한 힘	게임의 이익 주구의 의한 의사원장과 사유재산 제도를 근간으로 하는 자부존의 경제 질서는 한마다로 시장경제로 요약됩니다. 이런의 시장은 자유로운 의 시장에서 결정되는 가격은 생산자와 소비자 사이에서 정보를 전달하는 신호의 역할을 하여 자원배분이 효율적으로 이루어지도록 합니다. 생각해 보면 산년 이제는 생물활산품의 된 휴대폰을 예로 돌아봅시다. 같은 가능을 갖고 있는 경우 휴대폰의 환가에 큰 자이가 없을 것입니다. 그러나 전설에서는 모임에 따 또 다른 예를 들어 보겠습니다. 사과의 가격이 오르며 열반 사람들은 사과 소비를 들어고 그 대신 값이 성다적으로 오르지 않은 배약 같은 다른과 일을 더 어처럼 무엇을 생산했 것인지, 연락에 상실할 것인지, 또 생산된 재화를 어떻게 나누어야 할 것인지의 실제문제를 처음에서 영화된 누가리 해결해 줍니다. 한편 자본주의 사장경제와 대비되는 경계체제로 사회주의 계획 전체가 있습니다. 구소한이나 사장경제를 받아들이가 전의 중국, 배트를 들어 이 이해 태당됩 오늘날 세계는 자존주의 시장경제체계가 구류를 이루고 있습니다. 그렇지만 아느 나무는 수관 사장경제를 받아들이가 없음 나다. 사장경제가 되었다.

2차 전처리까지 마친 CSV 문서의 데이터 사이즈는 130 rows x 2 columns 입니다.