

인공지능 데이터 전처리 결과서

SKN Family AI Camp 2기 : 최종 프로젝트 2조



프로젝트 주제	경제 지식이 없는 국민들을 대상으로 경제 지식을 쉽게 설명하는 AI 인플루언서 웹 페이지 관리 Tool
사용 데이터	<p>[한국은행 제공 경제교육용 PDF 문서]</p> <ul style="list-style-type: none"> - 경제금융용어 700선 - 알기 쉬운 경제 이야기 (일반인용)
'경제금융용어 700선' 전처리 과정 (1)	<p>[원본]</p>  <p>'경제금융용어 700선'은 크게 용어 명칭, 용어 설명, 연관 검색어로 구성되어 있습니다.</p> <p>[사용 Library & Tool]</p> <p> LangChain</p> <p>PDF 속 텍스트 추출 시에 langchain_unstructured 라이브러리를 사용했습니다.</p> <p>해당 라이브러리는 pdf와 word 문서에서 text를 추출하는 데에 강점을 갖고 있습니다.</p> <p>또한 metadata 속성을 통해 문서 구조에 접근할 수 있다는 장점이 존재합니다.</p> <p> pandas</p> <p>추출한 텍스트는 판다스 라이브러리를 사용하여 CSV 문서 형태로 정리했습니다.</p> <p>용어 명칭, 용어 설명, 연관 검색어를 필드 형태로 가공한 후에 데이터프레임 형태로 정리했습니다.</p> <p>[1차 전처리 : PDF 속 텍스트 추출 (langchain_unstructured 라이브러리)]</p> <ol style="list-style-type: none"> 1) PDF 파일 내용을 텍스트로 변환 2) metadata 속성이 'Title'이거나 'NarrativeText'인 항목만 추출 3) 목차에 해당하는 자음 요소를 제거 <p>[1차 전처리 결과를 : 텍스트 파일]</p> <p>가계수지 가정에서 일정 기간의 수입(명목소득)과 지출을 비교해서 남았는지 모자랐는지를 표시한 것을 가계수지(household's total income and expenditure)라 한다. 가계수지가 흑자를 냈다면 그 가정은 벌어들인 수입 일부만을 사용했다는 것을 의미하며, 적자를 냈다면 수입 외에 빚을 추가로 얻어 사용한 것이라고 보아야 한다. 우리나라는 통계청에서 가계의 수입과 지출을 조사하여 국민의 소득수준 및 생활실태를 파악하기 위해 표본으로 선정된 가계에 가계부를 나누어 주고 한 달간의 소득과 지출을 기록하도록 한 다음 이를 토대로 가계수지 통계를 작성하여 발표하고 있다. 가계부의 소득항목에는 근로소득 사업소득 재산소득 이전소득 항목이 있고, 비용항목에는 식료품비 주거비 수도광열비 보건의료비 교육비 항목이 있다. 연관검색어 : 경상수지, 재정수지</p>

전처리 과정 (2)

3) 데이터프레임을 CSV 문서 형태로 저장

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

전처리 과정 (1)

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 중요됩니다. 여기에 경제적으로 윤택한 것도 행복한 삶의 여러 가지 조건 중 하나라고 할 수 있습니다. 현대의 물질문명에 대해 비판이 없는 것은 아니지만 경제적 필요가 행복의 한 요소인 것만은 틀림없습니다.

행복을 추구하는 우리 삶의 많은 부분은 경제적 의사결정과 밀접하게 관련되어 있습니다. 우리는 일상생활에서 의사결정을 계속하지 않을 수 없습니다. 예를 들어 시장에서 장을 볼 때 일반 채소를 살 것인지 조금 비싸더라도 유기농 채소를 살 것인지와 비롯해 택시를 탈까 버스를 탈까와 같이 대중교통을 이용할 때의 사소한 결정에서부터 대학진학 여부, 직업 선택, 주택 구입 등 그도 중요한 결정에 이르기까지 어떤 결정을 해야만 할 때가 헤아릴 수 없이 많습니다. 이러한 의사결정을 할 때 경제상식과 지식을 적절히 활용한 경제학 사고방식은 현명한 의사결정을 내리는 데 큰 도움이 됩니다.

그렇다고 경제지식이 사람들을 곧바로 부유하게 해주는 것은 아닙니다. 경제학 강의 첫 시간에 교수들이 강조하는 말은 “경제학이 부자가 될 수 있게 준 배는 방법을 가르치는 학문은 아니다. 라는 것입니다. 그렇다면 경제지식이 왜 필요할까요? 과연 이 책을 읽기 위해 투자한 시간의 가치에 상당하는 결실이 돌아올까요. 경제상식이나 지식이 많다고 해서 모든 사 일이 곧 부자가 되고 행복해질 수 있다면 참 좋을 것입니다. 그러나 현실은 그렇지 않습니다. 역사적으로 저명한 경제학자라고 해서 반드시 부자가 아 아니었던 것은 이것을 잘 대변해줍니다.

경제상식이나 지식이 무리를 한 순간에 부자로 만들어 주지는 않습니다. 그렇지만 무인이 일상사에서 합리적인 의사결정을 할수 있도록 도와줌으로써 장기적으로 보면 개인의 행복을 크게 해주는 역할을 합니다. 주식 투자를 예로 들어 보겠습니다. 경제지식이 단기적으로 주식으로 돈을 버는 기법을 가르쳐 주지는 않습니다. 그렇지만 위험을 분산하고 경제의 흐름을 이해할 수 있도록 함으로써 중장기적으로 손실을 최소화하고 나아가 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]

title	content
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 무 행복을 추구하는 우리 삶의 많은 부분은 경제적 의사결정과 밀접하게 관련되어 있습니다. 우리는 일상생활에서 의사결정을 계속하지 않을 수 없습니다. 예를 그렇다고 경제지식이 사람들을 곧바로 부유하게 해주는 것은 아닙니다. 경제학 강의 첫 시간과 교수들이 강조하는 말은 '경제학이 부자가 될 수 있게 준 배는 경제상식이나 지식이 우리를 한 순간에 부자로 만들어 주지는 않습니다. 그렇지만 우리가 일상사에서 합리적인 의사결정을 할수 있도록 도와줌으로써 장기
국민경제 운영 성과는 국민의 경제이해력에 비례	이러한 경제상식과 지식이 세계경제와 나라경제에서 여러 가지 경제현 상의 흐름과 실태 그 인과관계 등을 이해하는 데 많은 도움이 되는 것은 토달할 나위 일반적으로 사람들은 개인보다 정부가 더 장기적인 시각을 갖고 나라를 운영한다고 생각합니다. 이것은 꼭 정치인이나 공직자들이 일반인보다 훨씬 더 뛰어나 그래서 나라경제를 잘 운영할 수 있는 훌륭한 정치인과 공직자를 선출하 는 것이 중요하여 이를 위해서는 경제에 대한 올바른 이해력을 갖춘 시민이 요구됨 국민의 뜨거운 관심을 받는 활동한 국가주요에서 좋은 성과를 내려면 경기를 뛰는 선수들의 역할은 물론이고 국가를 둘러싼 생태계의 역할이 뒷받침되어야 경제문제
경제활동은 선택의 연속 - 선택은 행복의 최대화 과정	우리는 경제생활에서 최소성 때문에 선택을 하며 살고 있는데, 모든 선택에는 항상 무엇인가를 포기해야 한다는 대가가 따릅니다. 즉 선택과 선택의 대가 예를 들어 집을 마련하 는 것을 포기해야 합니다. 주부가 생활비를 할 때도 마찬가지입니다. 돈을 가족의 건강을 위하여 많은 음식을 만드는 데 사용할 수! 모든 경제적 결정에는 대가를 지급해야 합니다. 따라서 최선의 선택을 하려면 각각의 다른 선택 대안에 대하여 이득과 손해, 편익과 비용을 따져 보고 서로 나 세상에 공짜 점심은 없다' 또는 선택이 잡으려다 잡히지 않더라'와 같은 말이 있는데 이는 기회비용의 의미를 잘 말해 줍니다. 우리는 선택을 할 때 어떤 선택 저쪽도 하나의 경제적인 선택입니다. 아려울 값을 대비하거나 폭탄을 만들거 위한 저축은 결과적으로 미래 소비생활에 쓰이게 됩니다. 따라서 저축은 미래
하나라를 가지면 다른 하나를 잃는 기회비용	앞에서 우리는 최소성 때문에 발생하는 경제문제를 해소하기 위하여 편익과 비용을 비교해 합리적인 선택을 해야 한다고 하였습니다. 여기에서는 이에 덧붙 다른 예를 들어 보겠습니다. 어느 항공사가 예약 값이 공항에서 대기하다가 비행기를 타려는 사람에게 항공료를 얼마 받아야 하는지 결정한다고 합시다. 30 위의 사례와와 같이 선택에 따르는 추가적인 행복과 추가적인 희생을 비교하여 선택하는 것이 경제적 사고방식에 의한 의사결정인 것입니다. 실제 생활에
한계적으로 생각하여야 합리적인 의사결정	한편 경제주체들이 편익과 비용 개념을 한계적으로 고려하여 의사결정을 할 때, 주변 여건이 바뀌면 의사결정에 영향을 받습니다. 이처럼 의사결정에 영향을 미친 경제에 있어서도 마찬가지입니다. 늘어나는 최저비 지출을 위해 재정확자가 누적으로 정부가 재정 균형을 위해 소득세와 법인세의 세율을 인상하는 등 따라서 선택하는 데 중요한 고려요소가 되는 각종 규제, 조세제도 등의 유인을 변화시킬 때에는 이 변화에 대한 경제주체의 반응을 미리 염두에 두어야 하니
유인이 의사결정에 미치는 영향은?	수요는 소비자가 어떤 상품을 일정한 가격을 치르고 사려고 하는 의도를 말합니다. 공급은 생산자가 어떤 가격에 판매하고자 하는 의도를 말합니다. 수요는 마찬가지로 공급은 이윤을 최대화하려는 생산자의 합리적 의사결정 결과이며 상품 가격과 생산비용 등의 움직임에 따라 변합니다. 일반적으로 각기 다른 통 이와 같이 수요와 공급은 경제현상을 이해하는 데 필요한 아주 기본적인 개념인 동시에 꼭 알아야 하는 중요한 분석수단입니다.
수요와 공급은 모든 경제현상을 이해하는 필수 개념	경제문제는 크게 미시경제와 거시경제의 두 측면에서 분석할 수 있습니다. 가계나 기업의 의사결정 등과 같이 개별 경제주체의 관점에서 경제행위를 분석하 어떤 산의 특성을 파악하기 위해서는 손과 나무의 모습을 모두 알아야 하듯이 이들 분야를 모두 잘 이해할 필요가 있습니다. 마찬가지로 나라경제 제도 서로 2 개인이 어떤 연구에 의한 의사결정과 사유재산 제도를 근간으로 하는 자본주의 경제 질서는 한미대로 시장경제로 요약됩니다. 이러한 시장요 자유로운 의사
미시경제는 나무, 거시경제는 숲	시장에서 결정되는 가격은 생산자와 소비자 사이에서 정보를 전달하는 신호의 역할을 하며 자질배분이 효율적으로 이루어지도록 합니다. 상대적으로 보면 신비 이제는 생활불능수치에 한 집단쯤 예로 들어봅시다. 같은 기능을 갖고 있는 경우 집단들의 한가이 큰 차이가 있을 것입니다. 그러나 현실에서는 모멸에 따라 또 다른 예를 들어 보겠습니다. 사과의 가격이 오르면 일반사람들은 사과 소비를 줄이고 그 대신 값이 상대적으로 오르므로 사과와 같은 사과와 애플을 더 소 이처럼 두것을 생산할 것인지, 어떻게 생산할 것인지, 또 생산한 재화를 어떻게 나누어야 할 것인지를 경제문제를 시장에서 형성되는 가격이 해결해 줍니다. 한편 자본주의 시장경제와 대비되는 경제체제로 사회주의 계획경제가 있습니다. 구소련이나 시장경제를 받아들이기 전의 중국, 베트남 등이 이에 해당되나 오늘날 세계는 자본주의 시장경제체제가 주류를 이루고 있습니다. 그렇 지만 어느 나라도 순수한 시장경제를 채택하고 있지는 않습니다. 시장경제가 완벽하
교환의 이익과 가격의 신비한 힘	

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