

<Prediction Of GDP Per Capita By Indirect Indicators>

201300938 최영조

Description :

전 세계 국가들의 각종 지표를 데이터로 삼고 그것을 바탕으로 1인당 GDP를 예측할 수 있는 모델을 구축한다.

수집된 지표들은 지리, 정치, 사회 등등 다양한 분야의 통계로서 경제지표인 1인당 GDP의 산출에 직접적인 계산 요소로 포함되지는 않지만, 간접적인 상호작용을 통해 경제 기반에 분명한 영향이 있음을 명시적으로 보여주는데 이 분석의 1차적 목적이 있다.

예측된 순위를 바탕으로 원하는 만큼의 지표를 가진 가상국가의 1인당 GDP의 값 역시 예측 가능하다.

OECD국가들이나 선진국들의 평균 통계치들을 구하고 대한민국이 그에 미치지 못하는 지표는 어느 것인지 파악한다. 그 지표들을 개선했을 때 대한민국의 1인당 GDP는 어느정도 선까지 올라갈 수 있을지 예측해볼 수 있다. 혹은 평균 이상의 지표들이 하락했을 때 인당 GDP에 미치는 영향이 클지 예측해볼 수 있다. 이러한 분석들을 바탕으로 하여 한국경제발전의 방향성을 구체화하는데 이 분석의 최종적인 목적이 있다.

Plan

1. 웹상에서 1인당 GDP와 필요한 지표들을 스크래핑하여 합쳐진 데이터로 만든다.
 - 1-1. merge과정에서 항목별로 국가명 표기가 일치하지 않아 수정을 병행해야 한다.
이때는 국제 사회에 관한 약간의 배경지식을 활용한다.
Ex> '중국'/'중화인민공화국', '스와질란드'/'에스와티니', '북마케도니아'/'마케도니아'
 - 1-2. 긴 시간이 걸릴 수 있고 중간에 컴퓨터가 예기치 못한 에러를 일으킬 가능성도 다분하므로 과정마다 엑셀파일을 저장하며 진행한다.
2. 완성된 데이터는 전처리 과정을 통해 컴퓨터가 인식할 수 있는 상태로 만든다.
 - 2-1. 1인당 GDP와 각 지표마다의 연관관계의 시각화를 통해 빈 값을 어떻게 채울지에 대한 insight를 얻는다.
 - 2-2. 지표간 절대치의 크기차가 너무 심하면 정규화, 표준화 혹은 기타 스케일링 과정을 통해 전처리한다.
3. 머신러닝을 활용해 GDP 예측에 높은 정확도가 나오는 모델을 구축한다.
4. OECD 국가들의 각 지표의 평균치를 이용해 대한민국 사회 제반을 의미 있게 비교한다.
5. 안 좋은 지표가 개선되었을 경우 혹은 좋은 지표가 악화되었을 경우 대한민국의 1인당 GDP가 어느 위치에 있을지 예측해본다.

고려할 지표(data source) :

대부분의 지표는 위키피디아에서 얻어올 수 있다. 단, 위키백과 내에서도 출처가 확실한 경우만을 사용한다. 아래 지표들은 위키백과에서 인용한 1차 출처에 대해서만 표시하였다. 위키백과 밖에서 얻은 자료는 설명으로 적어놓았다.

- 1인당 GDP : 국제통화기금(2019년)
기준 국가들은 1인당 GDP 자료를 구할 수 있는 국가 중 UN 가입국으로 한다.
- 인지도 : 구글에 국가명을 영문으로 검색했을 때 나오는 결과 수를 기준으로 한다.
- 면적 : CIA 월드팩트북(2005년), '면적' 항목에는 출처가 없으나 '해안선' 항목에 표시돼 있다.
- 인구 : CIA(2010년)
- 평균수명 : UN(2018년)
- 출산율 : CIA 월드팩트북(2018년)
- 자살률 : WHO(2017년)
- 실업률 : 국가별로 취합한 기관과 년도가 약간씩 다르지만 위키피디아 '실업률에 따른 나라 목록' 항목을 보면 출처를 일일이 확인할 수 있다.
- 문해율 : CIA 월드팩트북(2013년)
- 에이즈감염자수 : CIA 월드팩트북(2016~2018년)
- 위치지역 : 네이버백과 '유엔의 지역그룹' 항목에는 외교부 제공 UN국들의 지역별 그룹이 나와있다. 단 미국, 캐나다, 이스라엘 등은 서유럽권과 같이 묶여있다.
- 언론자유 : RSF(2019년)
- 민주주의지수 : The Economist(2017년)
- 부패지수 : 국제투명성기구(TI)(2018년)
- 정치체제 : 위키백과 '공산국가', '전제군주제' 항목
- 징병제 여부 : 위키백과 '징병제' 항목
- 석유매장 : U.S. Energy Information Administration(2016년)
- 내륙국 여부 : 해안선 길이를 전체 면적으로 나눈다. CIA 월드팩트북(2005년)
- 최상위 교육기관 보유 : '글로벌대학리더포럼' 참가 26개 대학의 소속 국가들
-

(징병제와 정치체제는 신뢰있는 조사기관이 굳이 필요하지 않아 위키피디아의 항목만 적어놓았다.)