

세계 위험도 평가

학번: 2018022

이름: 박태형

Github address: https://github.com/2018022-Park-taehyoung/machine_learning.git

1. 안전 관련 머신러닝 모델 개발의 목적

독립변수: WRI

종속변수: Exposure, Vulnerability, Susceptibility, Lack of Coping Capabilities, Lack of Adaptive Capabilities

WRI(World Risk Score of the Region): 세계 지역 위험도

Exposure: 자연 재해에 대한 위험 또는 노출도

Vulnerability: 경제적인 프레임워크에 따른 취약성

Susceptibility: 경제적인 프레임워크에 따라 영향을 받기 쉬운 정도

Lack of Coping Capabilities: 사회적 및 물질적 보안에 따른 대처 능력 부족

Lack of Adaptive Capabilities: 기후변화, 자연재해 등에 대한 적응 능력

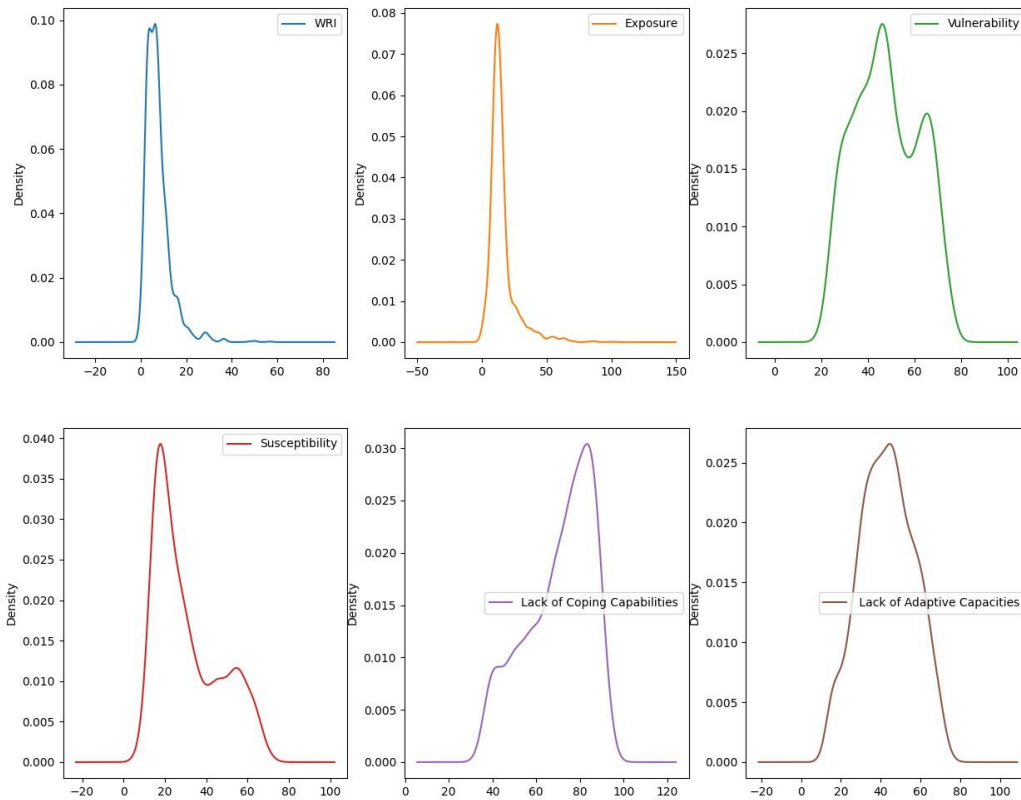
노출도, 취약성, 영향을 받기 쉬운 정도, 대처 능력, 적응 능력에 따른 세계 지역 위험도가 얼마나 되는지를 예측하는 머신러닝이다. 이 머신러닝으로 세계 지역 위험도를 평가하여 지역에서 스스로 재해에 대한 대책을 세움으로써 위험도를 낮출 수 있도록 한다.

2. 안전 관련 머신러닝 모델의 네이밍의 의미

세계 위험도를 평가하여 스스로 위험도를 낮추기 위해 노력하는 것이 목적이기 때문에 '세계 위험도 평가'라고 네이밍 하였다.

3. 개발 계획

먼저 독립변수와 종속변수로 해당하지 않은 데이터를 삭제한 뒤, 결측치를 삭제한다. 그 다음 밀집 그래프를 그린다.



머신러닝 모델은 선형 모델을 사용하고 여러 모델을 사용하고 mse 와 k-fold 를 사용해 학습 모델을 평가한다.

4. 개발 과정

먼저 데이터를 불러오고 독립변수, 종속변수로 사용하지 않는 데이터는 삭제한다. 데이터 안에 있는 결측치를 확인 후 삭제하고 밀도 그래프를 그려 시각화한다.

변수를 먼저 설정한 뒤 학습 모델을 선정해 학습시킨다. 이 때, 분류 모델을 사용했는데 오류가 나고 선형 회귀 모델을 사용해 학습시켰다.

제일 먼저 의사결정나무 모델을 사용했을 때의 mse 와 k-fold 값이다.

```
Mean Squared Error: 114.72768694516972
-127.75170332215313
```

그 다음 선형 회귀 모델을 사용했을 때,

```
Mean Squared Error: 81.32914751037659  
-100.77730411007914
```

릿지 회귀를 사용했을 때,

```
Mean Squared Error: 81.15166106142293  
-100.73050349069324
```

라쏘 회귀를 사용했을 때,

```
Mean Squared Error: 81.20534781346467  
-100.78281412582486
```

엘라스틱 넷 회귀를 사용했을 때,

```
Mean Squared Error: 81.16366951304497  
-100.74077868416597
```

각각의 모델의 성능을 평가했을 때, 의사결정나무 모델이 가장 떨어지고 나머지는 모두 비슷했지만, 릿지 회귀가 가장 성능이 좋다는 것을 확인할 수 있다.

5. 개발 후기

처음으로 데이터셋을 찾아보았는데 생각보다 많고 다양한 종류의 데이터가 있고 다른 사람들이 만든 코드를 보면서 참고를 할 수 있다는 점도 좋다고 생각했다.

선형 모델과 비선형 모델을 직접 해보고 오류를 찾으면서 어떤 차이점이 있는지 알게 되었다. 선형 모델에서 찾아보면서 여러 선형모델을 알게되었다.