**2021학년도 1학기 정보과학 프로젝트 제출 양식**

간단한, 수업 중 프로젝트 수행평가입니다.

1. 머신러닝,딥러닝을 활용해서 자기 관심 분야와 관련된 모델 구현하여 발표하기

**평가 관점은 다음과 같습니다.**

① 프로젝트에 대한 설명이 구체적으로 설명되어 있다.

② 프로젝트 작품이 충실하게 구현되어있다.

③ 과학 및 수학,정보 분야와 연관성을 가진 주제를 다루고 있다.

④ 과학 및 수학 분야에 대한 흥미와 호기심을 불러일으킬 수 있다.

⑤ 창의적 아이디어가 제시되어 있다.

⑥ 구체적 적용 상황이 제시되어 있다.

⑦ 과학 및 수학,정보 분야의 연구 활동에서의 활용 가능성이 높다.

**어떻게 만들어서 제출해야 하는가?**

- 다음 페이지의 양식에 맞추어 내용을 작성해 제출하면 됩니다.

- 수업시간을 충분히 활용해서 과제를 수행하세요. 미루고 있다가 집에 가서? 하는 것 금지!

- 매시간 끝날 때, 클래스룸에 **수업시간마다 제출합니다.**

- 매시간 프로젝트 완성도를 체크하여 세특 및 수행평가에 반영합니다.

1. 프로젝트 제출

- 파일 이름 : (1차시)2021정보과학프로젝트(학번,이름)

- 기본 제출

-(1차시)2021정보과학프로젝트(학번,이름)

- 머신러닝, 딥러닝 소스 파일

**이 양식 자료를 이쁘게?(글꼴, 배치 등) 꾸미려고 시간 버리지 말 것!**

**핵심으로 정확하게! 콕! , 분량은 중요하지 않지만, 노력은 중요함!**

**대표 팀:**

**학번** : 2130202 30209 30204

**이름** : 김다인 김진우 강민성

1-1. 머신러닝 or 딥러닝으로 000로 만들기 (몇번째 시간)

<https://www.kaggle.com/arashnic/earthquake-magnitude-damage-and-impact?select=csv_building_structure.csv>

참고문헌 https://ichi.pro/ko/gigye-hagseub-eul-tonghan-jijin-pihae-yecheug-4-bu-189272258882038

https://github.com/rivkms/atp\_ai/blob/main/main.ipynb

**제목 :** 건물의 특성과 지진 강도에 따른 건물 피해

**만든 것에 대한 설명**

- 관련 수학/과학 교과명 : 물리학, 지진학, 정보과학

- 관련 내용 또는 주제 : 건물의 특성(건물의 층수, 면적, 높이, 연도)과 지진 강도에 따른 피해 강도 (강도를 수치화, 피해 정도와 건물의 보수 정도)

1. 연구 주제 결정하기

|  |
| --- |
| 건물은 의식주를 해결하는 하나의 공간으로 안전이 보장되어야하는 중요한 공간이다. 따라서 건물을 화재, 재난, 자연재해로부터 사람들을 안전하게 지켜야 된다. 건물에 피해를 입힐 수 있는 대표적 자연재해로는 지진이 있다. 2016년 경주의 규모 5.8 지진 이후, 최근 우리나라에도 지진이 활발하게 발생하면서 대한민국도 지진의 안전지대에서 벗어날 수 없다는 이야기가 나오고 있다. 기상청의 지진 관련 자료에 따르면 규모 2내외의 사람이 감지할 수 없지만 지진계에는 기록되는 지진들은 꾸준히 발생하며, 국내지진 발생추이를 알아보면 2020년에는 규모 2이상 지진이 68회 발생한 것으로 기록되었다. 특히 우리나라 건축물들은 내진설계가 되어있지 않은 경우도 많아 건물의 특성이 지진에 따른 피해에 영향을 끼치는 경우가 많다. 특히 우리나라 건축물들 중 내진설계가 법적으로 의무화 되기 이전인 1988년 이전에 건설된 아파트나 고층 건물들, 3층 미만의 건물들은 내진설계에 관련한 법이 제대로 적용되지 않고 있다. 그래서 본 연구에서는 지진이 건물에 어떤 피해를 얼마나 주는지에 대해 알아보고, 건물의 높이, 건물의 수명 등의 건물 별 특성을 통해 지진의 영향을 받았을 때 건물이 입을 피해를 예측해보고자 한다. |

1. 문제 분석하기

|  |
| --- |
| 본 연구에서는 kaggle을 통해서 획득한 2015년 4월 네팔 고르카 지방에 발생한 규모 7.8 지진에 대한 자료를 사용할 것이다. 여러 csv파일 중 건물의 특성(층수, 건물이 지어진 연도, 지붕과 바닥의 재질)에 따른 피해 유무와 피해정도, 복구정도를 알아볼 수 있는 csv-building-structure.csv 파일을 머신러닝에 이용하였다.  -각 건물의 특성들을 분석 및 조사하여 건물의 특성과 지진 강도의 관계를 살펴본다.  -건물의 특성과 피해정도에 관한 그래프를 통해 어떠한 특성이 피해정도에 많은 영향을 주는지 알아보고 가중치를 설정한다.  -분석한 자료를 종합하여 건물의 특성에 따른 피해정도를 예측하는 모델을 만든다. |

1. 관련 내용 학습하기

|  |
| --- |
| 1. 머신러닝   머신 러닝의 종류 중 지도학습은 훈련 데이터로부터 하나의 함수를 유추해내기 위한 방법으로 주어진 데이터가 어떤 종류의 값인지 표식하는 것을 분류 그리고 유추된 함수 중 연속적인 값을 출력하는 것을 회귀라고 한다. 회귀 분석에서도 하나의 종속변수와 여러 독립변수 사이의 관계를 규명하고자 할 경우 이용되는 다중회귀분석을 본 연구에서 이용하고자 한다.   1. 지진   지진은 화산의 활동이나 단층, 함몰 등에 의해 대지가 진동하는 것이다. 지진 내부의 어떤 곳에서 어느 정도 이상의 급격한 움직임이 일어나 그 곳으로부터 지진파를 발하여 지표까지 전해진 현상을 이룬다. 진도는 지진의 크기를 나타내는 척도로 나라마다 다른 규정을 사용한다. 우리나라는 12단계로 분류되어 있는 수정 메스칼리 진도 계급(MMI)를 사용하고 있다.   1. 건물   건물은 토지에 정착된 구조물을 뜻한다. 법적으로 지붕, 기둥, 벽 이 세가지가 필수요소이며, 지하 또한 건물의 한 부분으로 친다. 각 건물은 고유의 건물번호를 지정받게 되고 건물이 위치한 지역의 번호, 층수, 높이, 지붕의 재질, 주추의 유무 및 재질 등으로 건물을 분류할 수 있다. |

4. 데이터셋 획득과정과 전처리 과정을 구체적으로 제시하기

|  |
| --- |
| 해당 데이터셋은 캐글의 arashnic/earthquake-magnitude-damage-and-impact에서 얻을 수 있었다. 해당 데이터셋 중에서 csv\_building\_structure.csv파일을 사용하여 이번 인공지능 제작에 사용하였다. 해당 데이터를 pandas를 이용하여 불러온 후 다음과 같은 전처리를 실행하였다.  지진에 대한 피해에 관련성이 없는 column들인 building id, district id, vdcmun\_id, ward\_id, count\_floors\_pre\_eq, count\_floors\_post\_eq, height\_ft\_post\_eq, roof\_type, plan\_configuration, has\_superstructure 관련 column, technical\_solution\_proposed은 필요 없으므로 이와 관련된 columm을 del이라는 하나의 배열에 저장하는 방식을 차용하여 이 배열에 저장된 columm들을 드랍시킨다. 이후 plan\_configuration에 1개, position에 1개, damage\_grade에 12개로 총 14개의 결측치가 있는 것을 확인하였고, 크게 머신러닝에 영향을 미치지 않을것이라 예상하여 dropna함수를 이용하여 해당 row를 드랍시켰다. land\_surface\_condition, foundation\_type, ground\_floor\_type, other\_floor\_type, position, damage\_grade 의 6개의 columm에서 object type인것을 확인 하였고, damage\_grade는 해당 정도를 숫자로 변환하고, 나머지는 원핫인코딩을 통하여 전부 int형으로 바꾸었다.  해당 데이터셋의 전처리와 밑에서 나오는 두가지 방법의 모델에 공통적으로 사용하였고, 이를 프로그램 코드상 data3이라는 변수에 저장해두었다. data3의 데이터들의 상관관계는 다음과 같다. |

5. 모델 설계하기(팀원들끼리 분담하여 작성해봅시다)

|  |
| --- |
| C언어를 이용하여 직접 작성해보자  <https://github.com/rivkms/atp_ai/blob/main/main.ipynb>  프로그램 코드는 위 깃헙 사이트에 업로드 해두었다.  모델은 총 두가지를 설계하였다. 단순 회귀를 이용하는 방법(i)과 딥러닝 회귀 방법을 이용하는 인공지능(ii)이다.  i의 경우 damage\_grade를 종속변수로, 나머지 columm들을 독립변수로 설정하였다. 해당 모델은 LinearRegression를 이용하였고, test set의 크기는 0.3으로 설정하였다.    ii의 경우도 i과 같은 방식으로 종속변수랑 독립변수를 설정한다. StandardScaler를 이용하여 x와 y를 세팅한 후 train set과 test set으로 나누었다. 여기서 test set의 크기는 0.2를 사용하였다. 이 후 Sequential를 이용하여 Dense가 50인 relu 출력층을 만들었고, 이후 Dense가 180인 relu 은닉층 2개, Dense가 50인 linear 은닉층 하나를 만든 후 이를 출력층으로 내보냈다. 이 각 은닉층 과정 후 0.3의 데이터셋의 Dropout이 이루어졌다.  그림 1 model summary  이후 해당 모델을 fit 시킨다. 이 때 epoch는 50으로 batch\_size는 20으로 설정한다. |

6.모델 학습 후 성능 파악하기

|  |
| --- |
| 완성된 프로그램을 실행하여 다양하게 입력하고, 출력값을 산출하여 정확성 여부를 따져보자  i에 대하여 모델의 성능을 파악하였다. MSE, RMSE, MAE 세 값을 이용하여 해당 모델을 평가하였다. 해당 모델에 대한 평가는 다음과 같이 나왔다.    ii에 대하여도 모델의 성능을 파악한다. 모델의 학습데이터의 손실률을 그래프로 확인하여 보았을 때 다음과 같다.  해당 그래프를 통하여 약 3~4회 정도 까지 train set의 손실률이 매우 급격히 떨어지는 것을 알 수 있었고, 이후에는 비슷한 손실률을 가지는 것으로 파악되었다. 이 또한 마찬가지로 MSE, RMSE, MAE 세 값을 이용하여 해당 모델을 평가하였다. 이 때 해당 모델의 평과는 다음과 같이 나왔다. |

7.모델 공유 및 활용 방법 구상하기

|  |
| --- |
| 현재 연구에 사용한 데이터가 네팔이라는 한 나라에서만 측정된 데이터이므로 다양한 지반이 속한 나라들을 대상으로 하여 지진의 규모와 건물의 특성에 대한 더 많은 자료들을 수집한다면 모든 나라들을 대상으로 일반화하거나 나라별 특징을 파악할 수 있으므로 보다 안전한 건축물을 제작하는 데에 있어 기준이 되며 많은 도움을 줄 수 있을 것이라고 생각한다.  또한 현재 구축되어있는 건축물에 대하여 지진이 발생하였을 때 해당 건물의 안정성을 파악할 수 있는 하나의 기준으로서 사용할 수 있을것이다. 미리 지진이 발생하였을 때 어떤 건물이 붕괴될지를 파악할 수 있다는 것은 미리 대피책을 마련한다던지, 건물의 보완설계 등을 통하여 건물의 안전성을 강화시킬 수 있는 대책을 마련할 수 있는 계기가 될 수 있을 것이다. |

8.(중요)프로그램을 제작하며 새로 알게 된점, 추가로 넣고 싶은 기능등을 작성해보자

|  |
| --- |
| 대한민국의 경우에는 건물의 위험 등급을 A, B, C, D, E등급으로 나누게 되는데 나라마다 건물의 특성이 다 다르기에 이 등급에 맞게 대한민국에서의 자료를 수집하여 연구를 진행해본다면 내진설계와 보다 안전한 건물들을 설계 및 제작하는데에 유용한 연구가 될 수 있겠다는 생각이 들었다.  지진의 규모는 선형함수가 아니라 log함수 꼴로 나타나지기 때문에 단순히 개연성을 찾을 수는 없어 다양한 규모에 대해 파악해보아야하는데, 본 연구에서 규모 7.8인 지진에 대해서만 예측을 진행한 점이 아쉬웠다. 규모에 따른 건물의 피해를 알아본다면 조금 더 활용도가 높은 연구가 될 수 있을 것이라는 생각을 하였다.  본 연구에서 사용된 자료들은 대부분 10피트정도의 저층 건물에 대한 자료였지만, 우리나라에서 적용하기 위해서는 고층 건물에 대한 자료가 더 필요하다.  이번 연구에서는 네팔의 지진에 대해서 네팔의 건물들의 붕괴 정도를 데이터 셋으로 잡았다. 하지만 지진별로, 그 세기가 다르고, 각 건물의 지형 무름정도 혹은 건물 양식 및 건설방식에 따라 그 건물의 붕괴 정도는 전부 다르게 나올 것이다. 추가적으로 이 프로그램을 제작하여 상용화를 하기 위해서는 많은 지진들에서, 각 지진들의 세기와 각 건물별 피해정도가 상세하게 정리되어야 할 것이다. 또한 각 건물의 사용 재료 등에 대해서도 또한 더 구체적으로 정리하여야 한다. 이런 데이터셋을 수집하여 다시 한번 해당 모델을 설계하여, 전세계의 지진의 세기별 및 각종 데이터를 바탕으로 해당 건물의 지진에 대한 안정성을 예측하는 프로그램을 만들어보고싶다. 또한 이 때 회귀에서 lineRegression만을 사용하였는데 polynomialRegression등 다양한 방법으로 만들어 본다면 더 좋은 결과값이 나올 수 있을지 한번 알아보고 싶다.  이번 연구를 통하여 데이터들의 전처리 과정에서 데이터 정보를 확인하고 필요없을것이라 예측되는 데이터를 날리는 과정에서 어떤 데이터가 의미없는 데이터일지 파악하는 방법에 대해 친구들과 토의를 통하여 알아볼 수 있었고, 여러가지 방법으로 회귀 모델을 만드는지, 각 모델에서 어떤 점을 변화시키면서 더 좋은 성능을 알아낼 수 있는지를 알게 된 것 같다.  (Grade 1, Grade 2: No need, Minor repair // Grade 3, Grade 4: Major repair // Grade 5: Reconstruction) |