# **[양식1] 2021 데이터사이언스 ‘나만의 프로젝트’ 계획서**

|  |  |  |
| --- | --- | --- |
| 항목  Topic | 내용  Content | 비고  Note |
| 팀이름  Team Name | 딜러 |  |
| 팀원이름  Team Members | 201600282 엄기산  201600294 임승호 |  |
| 제목  Title | 중고차 거래 데이터셋을 통한  중고차 시세 예측 |  |
| 목적 (프로젝트로 해결하고 싶은 문제)  Objective (what you want to solve) | 중고차를 사거나 판매를 할 때 보다 정확한 시세를 알고 현명한 거래를 할 수 있도록 도와준다. |  |
| 프로젝트 요약  Project Summary | 거래할 중고차의 정보를 넣어 예상 가격을 측정한다. |  |
| 모델 / 입력 / 출력  Model / Input / Output | 모델 : ANN  입력 : 차종, 연도, 주행거리, 유종, 기어, 연비  출력 : 가격 |  |
| 장점 / 독창성  Advantages / Originality | 장점 : 현명한 거래를 도와준다.  독창성 : 한국의 데이터가 없어 다른 나라의 데이터를 한국의 단위로 환산하여 신경망을 구축하였다. |  |
| 구현 기술 Python / PyTorch  Implementation Technology | 구현기술 : Python & PyTorch |  |
| 데이터 확보 방법 / 데이터 특징 / 데이터 전처리  Dataset Acquisition / Data Characteristics / Data Preprocessing | 데이터 확보 방법 : Kaggle 검색  데이터 특징 : UK의 데이터셋  데이터 전처리 : UK의 단위로 이루어진 데이터들을 모두 한국에서 사용하는 단위로 환산하였다. 또한 문자열의 값들은 mapping을 통해 정수로 변환하였다. |  |

**2021 데이터사이언스 ‘나만의 프로젝트’ 보고서**

# **중고차 거래 데이터셋을 통한 중고차 시세 예측**

|  |  |
| --- | --- |
| 제1저자 엄 기산  소속 수학과  학번 201600282  yonghi7624r@inu.ac.kr | 제2저자 임 승호  소속 수학과  학번 201600294  ho4788@inu.ac.kr |

**요 약**

자동차의 모델명, 연식, 변속타입, 주행거리, 연료타입, 세금, 연비, 엔진타입 등과 같은 데이터들을 통해 ANN모델을 사용하여 중고차 가격을 예측한다. 이때 hidden 노드의 개수, learning\_rate, epoch를 수정해가며 하이퍼 파라미터를 찾는다.

# **서론**

중고차의 가격이 중고차 시장에 따라 딜러에 따라 너무나 천차만별인 가격으로 형성됩니다. 이러한 문제를 해결하기 위해 중고차들의 적정가격을 제공해 조금이라도 소비자가 현명한 소비를 할 수 있도록 도와준다.

# **관련연구**

아우디 중고차의 가격 예측 모델 연구를 참고하여 국내 현대차의 여러 모델에 대하여 중고가를 예측하였습니다. 이때 선형회귀로 구현 되어있던 자료를 저희가 새롭게 ANN모델로 바꾸는 과정을 수행하였습니다. 출처는 참고문헌에 기입하였습니다.

# **데이터**

영국의 중고차 데이터들을 가져와 국내 단위로 변형 시켰습니다. 또한 다양한 브랜드가 아닌 국내 브랜드인 현대 자동차의 중고차 데이터들을 모아왔습니다.

[여기에 데이터 맨처음에 보여주는 사진 캡쳐좀 해서 올려줘! ]

# **방법**

Tensor로변환하기위해서 그리고 nn모델에 넣어 train하기 위해 모든 데이터들을 숫자형태인지 확인할 필요가 있었다.

[ 여기에 model, transmission, fuelType의 타입 출력한 부분 캡쳐해서 올려줘 ]

먼저 확인 결과 model, transmission, fuelType이 object형식이어서 각 열에 따른 unique한 값들을 확인해 보았다.

[ unique 값들 캡쳐해서 올려줘 ]

그 결과 model 에서는 총 16가지, transmission과 fuelType은 4가지의 경우로 나뉘어서 각 문자열을 1~16 그리고 1~4 로 맵핑하는 과정을 수행하였다.

[ 맵핑하는거좀 캡쳐해서 올려줘 ]

또한 tax는 중고차 가격 측정에 많은 관련이 없다고 판단하여 tax는 drop함수를 통해 제거하였다.

[drop하는 부분좀 ! 올려줘 ]

이와 같은 데이터 전처리를 수행 한 후 ANN 모델을 사용하여 예측을 수행한다.

# **실험**

Hidden node의 개수와 learning rate, epoch를 조절하면서 최선의 결과를 찾아보았다.

[ 여기 아래부터 각 node와 러닝 레이트, epoch에 따른 정확도 맨아래부분을 출력해서 캡쳐좀 부탁할게 ]

1. 실험 1( hidden node = , learning rate = , epoch= )

위의 방법에서는 정확도가 %나왔다.

1. 실험 2( hidden node = , learning rate = , epoch= )

위의 방법에서는 정확도가 %나왔다.

1. 실험 3( hidden node = , learning rate = , epoch= )

위의 방법에서는 정확도가 %나왔다.

1. 실험 4( hidden node = , learning rate = , epoch= )

위의 방법에서는 정확도가 %나왔다.

1. 실험 5( hidden node = , learning rate = , epoch= )

위의 방법에서는 정확도가 %나왔다.

# **결론**

Learning\_rate를 너무 많이 낮게 설정하게 되면 정확도가 30%정도 수준으로 너무 낮게 나왔고 최대한 높은 정확도를 보이고 싶었지만 우리가 수행해본 결과

대부분의 정확도는 50.2%정도로 나왔다. 그리고 최고의 정확도는 51%정도였다. 이렇게 낮은 정확도를 보인 이유는 먼저 price를 천만원 단위로 나눴기 때문에 이런 정확도를 보였다고 생각했고 또한 train 세트의 부족이라고 생각하였다.

# **참고문헌**

1. <https://www.kaggle.com/aryyawijaya/audi-used-cars-price-prediction>