

인공지능 과제

- Linear regression -

컴퓨터 정보통신 공학과 214683
장인환

목차

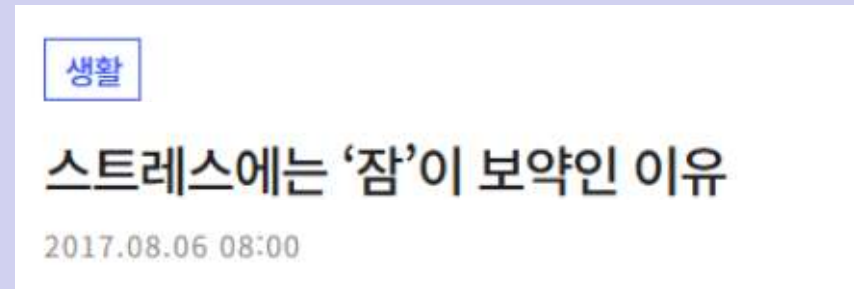
1. 모델링 배경 및 목적
2. 데이터 설명
3. 모델링 과정
4. 결과 및 성능



1. 모델링 배경 및 목적



모델링 배경 및 목적



최근들어 불면증의 초기증상이 보이게 되면서 잠에 대해 관심이 많아진 상태였습니다.

▲ 곧 시험기간이 다가오고 있는 만큼 스트레스를 받는상황을 줄이고자 데이터를 분석해보고자 합니다.

수면시간과 스트레스간의 관계를 알아보고 최적의 수면시간을 찾아보고자 하였습니다.

2. 데이터 설명



데이터 설명

```
sleep.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 374 entries, 0 to 373
```

```
Data columns (total 13 columns):
```

#	Column	Non-Null Count	Dtype
0	Person ID	374 non-null	int64
1	Gender	374 non-null	object
2	Age	374 non-null	int64
3	Occupation	374 non-null	object
4	Sleep Duration	374 non-null	float64
5	Quality of Sleep	374 non-null	int64
6	Physical Activity Level	374 non-null	int64
7	Stress Level	374 non-null	int64
8	BMI Category	374 non-null	object
9	Blood Pressure	374 non-null	object
10	Heart Rate	374 non-null	int64
11	Daily Steps	374 non-null	int64
12	Sleep Disorder	155 non-null	object

```
dtypes: float64(1), int64(7), object(5)
```

```
memory usage: 30.8+ KB
```

374개의 데이터셋

11개의 독립변수, 1개의 종속변수가 존재

종속변수 : Quality of Sleep(수면의 질)

Gender: 성별

Age: 나이

Occupation: 직업

Sleep Duration: 하루 평균 수면 시간

Quality of Sleep: 수면의 질

Physical Activity Level: 신체 활동 수준

Stress Level: 스트레스 수준

BMI Category: 체질량지수(BMI) 범주

Blood Pressure: 혈압

Heart Rate: 심박수

Daily Steps: 일일 걸음 수

Sleep Disorder: 수면 장애 여부

3. 모델링 과정



모델링 과정

1. CSV 파일을 통해 데이터 불러오기

```
[1]: import pandas as pd
sleep = pd.read_csv("Sleep time.csv")
```

```
[2]: sleep.head()
```

	Person ID	Gender	Age	Occupation	Sleep Duration	Quality of Sleep	Physical Activity Level	Stress Level	BMI Category	Blood Pressure	Heart Rate	Daily Steps	Sleep Disorder
0	1	Male	27	Software Engineer	6.1	6	42	6	Overweight	126/83	77	4200	NaN
1	2	Male	28	Doctor	6.2	6	60	8	Normal	125/80	75	10000	NaN
2	3	Male	28	Doctor	6.2	6	60	8	Normal	125/80	75	10000	NaN
3	4	Male	28	Sales Representative	5.9	4	30	8	Obese	140/90	85	3000	Sleep Apnea
4	5	Male	28	Sales Representative	5.9	4	30	8	Obese	140/90	85	3000	Sleep Apnea

2. 데이터 분리

```
[7]: y=sleep['Quality of Sleep']
x=sleep['Stress Level']
```

```
[11]: from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x,y,train_size=0.7, random_state=10)
```


모델링 과정

3. 데이터 차원 맞추기

```
[15]: type(x_train)
```

```
[15]: pandas.core.frame.DataFrame
```

4. 모델 선택 및 학습

```
[16]: from sklearn.linear_model import LinearRegression  
model = LinearRegression()  
model.fit(x_train, y_train)
```

```
[16]: * LinearRegression
```

```
LinearRegression()
```

4. 결과 및 성능



결과 및 성능

1. 예측값

```
[19]: y_pred=model.predict(x_test)  
y_pred
```

```
[19]: array([[6.33587894, 5.72930299, 8.1556068 , 8.76218275, 8.1556068 ,  
        6.33587894, 6.94245489, 7.54903085, 5.72930299, 7.54903085,  
        8.1556068 , 5.72930299, 6.33587894, 7.54903085, 8.76218275,  
        7.54903085, 6.94245489, 7.54903085, 6.94245489, 8.1556068 ,  
        8.76218275, 7.54903085, 6.33587894, 6.94245489, 8.76218275,  
        8.76218275, 5.72930299, 8.76218275, 5.72930299, 8.1556068 ,  
        6.33587894, 6.33587894, 8.1556068 , 7.54903085, 8.1556068 ,  
        8.1556068 , 8.76218275, 8.1556068 , 7.54903085, 8.1556068 ,  
        8.76218275, 8.1556068 , 8.1556068 , 5.72930299, 8.76218275,  
        6.94245489, 8.1556068 , 8.76218275, 5.72930299, 5.72930299,  
        6.94245489, 8.76218275, 6.94245489, 5.72930299, 5.72930299,  
        7.54903085, 6.94245489, 7.54903085, 7.54903085, 5.72930299,  
        6.94245489, 6.94245489, 8.1556068 , 7.54903085, 6.33587894,  
        5.72930299, 5.72930299, 8.76218275, 8.1556068 , 7.54903085,  
        6.33587894, 8.76218275, 6.94245489, 6.33587894, 6.33587894,  
        8.1556068 , 6.94245489, 5.72930299, 5.72930299, 7.54903085,  
        7.54903085, 7.54903085, 8.1556068 , 6.33587894, 6.94245489,  
        7.54903085, 5.72930299, 7.54903085, 5.72930299, 7.54903085,  
        6.33587894, 8.1556068 , 7.54903085, 6.94245489, 8.1556068 ,  
        8.76218275, 7.54903085, 6.94245489, 6.94245489, 6.33587894,  
        6.94245489, 7.54903085, 8.1556068 , 8.1556068 , 8.1556068 ,  
        8.1556068 , 6.33587894, 8.76218275, 8.76218275, 8.76218275,  
        7.54903085, 7.54903085, 8.1556068 ]])
```

2. 성능 평가

```
[28]: print("평균 절대 오차 MAE : ", mean_absolute_error(y_test, y_pred))  
print("평균 절대값 백분율 오차 : ", mean_absolute_percentage_error(y_t  
print("평균 제곱 오차 : ", mean_squared_error(y_test, y_pred))
```

평균 절대 오차 MAE : 0.3992173700368052

평균 절대값 백분율 오차 : 0.058362842753011276

평균 제곱 오차 : 0.2775805849628025

참고 자료

- <https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset>
- <https://www.sleep.or.kr/html/?pmode=BBBS0019600010&smode=view&seq=918>