*import* pandas *as* pd

*from* sklearn.preprocessing *import* MinMaxScaler

*import* numpy *as* np

*# 0. 데이터 로드*

df = pd.read\_csv('diabetes.csv')

*# 1. 결측치 처리. 0은 결측치로 간주하여 평균치로 대체*

columns\_missing = ['Glucose', 'BloodPressure', 'SkinThickness', 'Insulin', 'BMI' ]

*for* column *in* columns\_missing:

df.loc[df[column]==0,[column]] = np.nan

df[column]= df[column].fillna(df[column].mean())

*# 2. 이상치 처리. 상위 1%는 평균 값으로 대체*

outliers = {'SkinThickness': df['SkinThickness'].quantile(0.99), 'Insulin': df['Insulin'].quantile(0.99)}

*for* (column, outlier) *in* outliers.items():

df.loc[df[column]>=outlier, [column]] = df[column].mean()

*# 3. Age 열 스케일링*

minmaxscaler = MinMaxScaler()

df['Age'] = minmaxscaler.fit\_transform(df[['Age']])

*# 4. 결측치 개수 출력.*

*for* column *in* df.columns:

*if* column in columns\_missing:

missing = (df[column] == 0).sum()

print(f'{column}의 결측치 개수: {missing}개')

*else*:

missing = df[column].isnull().sum()

print(f'{column}의 결측치 개수: {missing}개')

*# 5. 상위 5개 행 출력*

print('\*'\*20)

print(df.head())

