The Rate of Obesity and Wealth Distribution/Growth

Phil Lyons

Machine Learning, Comp. Sci. - NWMSU

Intro/Goals

For the ML Project for Spring 2021 I held the question of if and how Obesity is affected by the wealth and poverty in the area. The source dataset I pulled is: https://www.Kaggle.com/annedunn/obesity-and-gdp-rates-from-50-states-in-20142017. This covers all 50 states from 2014-2017.

Dataset Features

The primary features are:

Adult.Obesity.Percent – mine

Average.Income (State avg.)

Poverty.Rate (Pov growth)

Adult.Obesity (avg. obesity)

Real.GDP.Growth (rate of inc.)

Exploration of the Data

With running data over the linear regression I had Avg.Income, Pov.Rate for x and Adult.Obesity for y. There were issues running through the train set, with it split 85/15 earlier, which produced an R2 score of .30 and .37. However, the test set produced .74 for R2 with a drastic increase.

```
Bias:  0.35963142174894813
Coefficients:  [-3.1272043e-06  1.0144648e-01]
MSE:  0.00067061574890684 02
r2:  0.3768515213676351
```

```
Bias:  0.26923357412799964
Coefficients:  [-2.14596160e-06  6.33036986e-01]
MSE:  0.00027266677578537764
r2:  0.7466330208003304
```

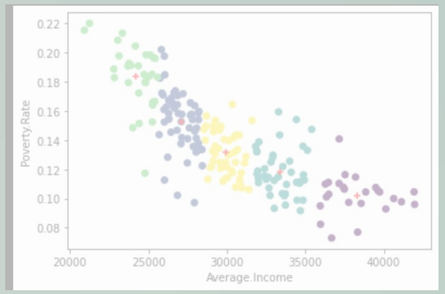
With the DT and Linear SVC models DT did the common thing of overfitting in the train set where Linear SVC had the values of .20 and .127 for accuracy and f1. However, when it came to the test sets as shown in the images, the results were very poor. DT having .15 for accuracy and .135 for F1 where Linear SVC had .066 for accuracy and .035 for f1 proving that these are not the features to be working with in the dataset.

```
DECISION TREE:

[[0 0 0 0 0 0 2 0 0 0 0 0 0 0]
 [0 0 0 1 0 0 0 0 1 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 1]
 [1 0 1 3 1 0 1 1 0 0 0 0 0 0]
 [0 0 1 1 0 0 1 0 0 0 0 0 0 0]
 [0 0 0 0 0 1 0 0 2 1 0 0 0 0]
 [0 0 0 2 0 0 2 2 0 0 0 0 0 0]
 [1 0 1 0 0 1 4 0 0 1 2 0 0 0]
 [0 0 0 1 0 1 0 0 1 0 1 0 0 0]
 [0 0 0 1 0 0 0 0 0 0 0 1 0 0]
 [0 0 0 3 0 1 1 1 0 0 0 1 0 2]
 [0 0 0 0 1 0 0 0 0 0 0 1 1 2]
 [0 0 0 0 0 0 0 1 0 0 0 1 1 0]
 [0 0 1 0 0 0 0 0 0 0 0 0 0 0]]
Accuracy:     0.15
Precision:    0.13068181818181818
Sensitivity:  0.15
F1:           0.13538126361655772
```

```
LINEAR SVC:

[[0 0 0 0 0 0 2 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 2 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 1 0]
 [0 0 0 0 0 0 5 0 0 1 0 0 0 2 0]
 [0 0 0 0 0 0 1 0 0 2 0 0 0 0 0]
 [0 0 0 2 0 0 2 0 0 0 0 0 0 0 0]
 [0 0 0 2 0 0 2 0 0 0 0 1 0 1 0]
 [0 0 0 1 0 0 6 0 0 3 0 0 0 0 0]
 [0 0 0 1 0 0 1 0 0 2 0 0 0 0 0]
 [0 0 0 0 0 0 1 0 0 1 0 0 0 0 0]
 [0 0 0 4 0 0 0 0 0 3 0 1 0 1 0]
 [0 0 0 0 0 0 0 0 0 1 0 1 0 3 0]
 [0 0 0 0 0 0 0 0 0 0 0 1 0 1 1]
 [0 0 0 0 0 0 0 0 0 0 0 1 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]]
Accuracy:     0.06666666666666667
Precision:    0.02832167832167832
Sensitivity:  0.06666666666666667
F1:           0.035396825396825406
```

Narrative

The main thing to be pulled from this exploration is when it comes to one's health, wealth does not affect it massively. The govt. and community are able to help GDP growth without consequences from the health and obesity of individuals. As wealth is not a factor there may be another leading cause to the rise in obesity whether it is fast food restaurants or their area of living (urban/rural) but from these results we can take out money and overall wealth as shown from the DT, Kmeans, and SVC.




```
DECISION TREE:

[[0 0 0 0 0 0 2 0 0 0 0 0 0 0]
 [0 0 0 1 0 0 0 0 1 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 1]
 [1 0 1 3 1 0 1 1 0 0 0 0 0 0]
 [0 0 1 1 0 0 1 0 0 0 0 0 0 0]
 [0 0 0 0 0 1 0 0 2 1 0 0 0 0]
 [0 0 0 2 0 0 2 2 0 0 0 0 0 0]
 [1 0 1 0 0 1 4 0 0 1 2 0 0 0]
 [0 0 0 1 0 1 0 0 1 0 1 0 0 0]
 [0 0 0 1 0 0 0 0 0 0 0 1 0 0]
 [0 0 0 3 0 1 1 1 0 0 0 1 0 2]
 [0 0 0 0 1 0 0 0 0 0 0 1 1 2]
 [0 0 0 0 0 0 0 1 0 0 0 1 1 0]
 [0 0 1 0 0 0 0 0 0 0 0 0 0 0]]
Accuracy:     0.15
Precision:    0.13068181818181818
Sensitivity:  0.15
F1:           0.13538126361655772
```

Cleaning and Dimension Reduction

Took out 3 features to help the run speed. They were copies of other features with *100 at the end. Used an RF later on Avg.Income, Pov.Rate – x; and Real.GDP.Growth – y; which produced some interesting results of having 0 for all values with acc, f1, prec, and sens. But with perfect matrix.

Future Directions

Dimension reduction needs to be relooked and to also find new potential causes.

Analysis

As mentioned before with the models and graphs used through this project it has shown there to be no relation of wealth and obesity whatsoever.