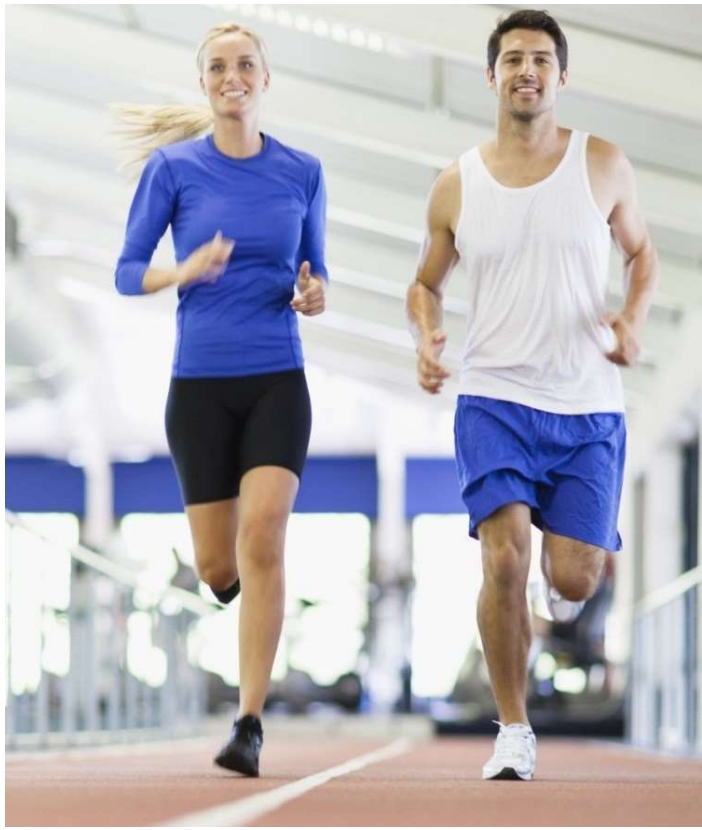


LIFESTYLE AND OBESITY

BY SAM LIM

BUSINESS PROBLEM

- Due to the pandemic, people's general lifestyles have changed
- As people are urged to stay indoors, their activity levels have decreased
- Decline in physical activity can lead to overweightness or obesity

GOAL

- By looking into variables such as family history of obesity, eating habits, and lifestyle habits predict a person's obesity level.
- Recommend a model to help predict a person's obesity level.
- Find which variables affect a person's obesity level the most

ST. JOSEPH COUNTY
DEPARTMENT OF HEALTH
Prevent. Promote. Protect.

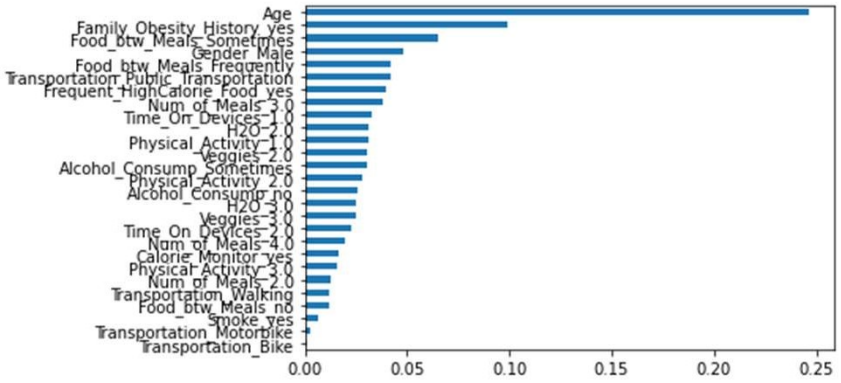
The South Bend Clinic

DATA AND MODELING

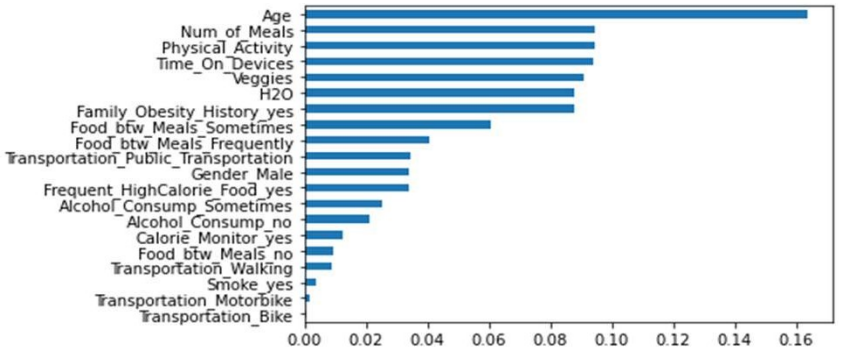
- Initially, a person's obesity level had 7 categories: underweight, normal weight, overweight 1 and 2, and obese level of 1, 2, and 3. Later, reduced to three: normal, overweight, and obese.
- Two variants of the same data. Because the initial data had higher numbers of people in the normal weight category compared to the sum of all other 6 categories, the data was changed so that the number of individuals in each category were more even. This process is known as SMOTE.
  - SMOTE'ing the data changed many of the categorical variables' values
  - One version of the data and models tried to view the changed data as categorical and the other viewed them as numerical.
- Random Forest, K Nearest Neighbors, and Logistic Regression models were used

WHAT ARE THE IMPORTANT FACTORS?

Random Forest Variable Importance (SMOTE)

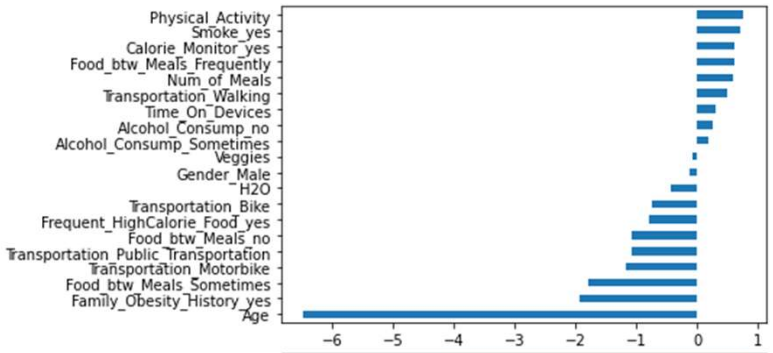


Random Forest Variable Importance (un-SMOTE)

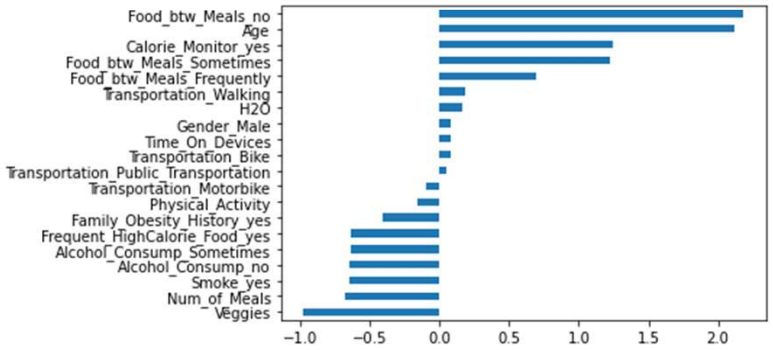


- When looking at the importance factors of the variables in the random forest model, we can see that age is the most important factor in both variations of the dataset.
- Other factors that we consider as healthy lifestyle habits are also shown to have positive effects on a person's obesity level

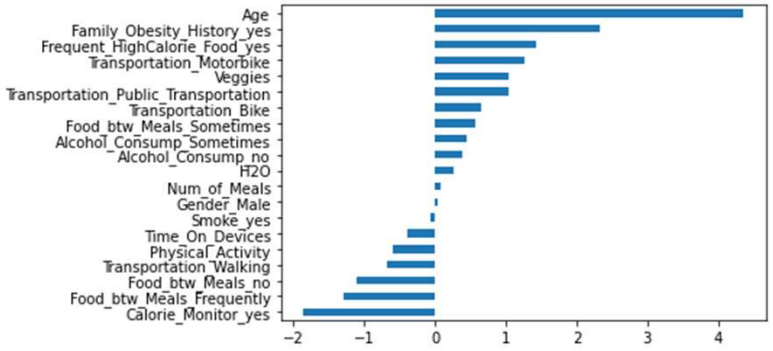
IMPORTANT FACTORS (CONT)



Logistic regression coefficient for normal weight category



Logistic regression coefficient for overweight category



Logistic regression coefficient for overweight category

- Using the logistic regression model, we can observe the coefficients of each variable for different obesity levels.
- For normal weight, age still affects the model the most even though it has a negative coefficient.

RECOMMENDATIONS

- Hydration, vegetable consumption, the number of meals, physical activity, reducing food consumption between meals, and eating less high caloric foods were the most effective in reducing obesity levels.
- As age was the biggest factor that contributed to obesity, the elderly should take extra care to stay in good shape.
- From our data, we can conclude that much of the habits that we view as healthy do benefit us.
- The best model to determine a person's obesity level is random forest as it was able to provide the best predictions.

CONCLUSION AND NOTES FOR FUTURE RESEARCH

- For future research, I believe that looking into some correlation between the different lifestyle choices and eating habits of a person could better represent how each variable affects a person's obesity level.
- Some uncontrollable factors such as family history of obesity and age seemed to play the biggest role in a person's obesity level.
- Another factor that we could investigate could be the correlation between a person's age and their lifestyle habits. It is known that the elderly often do not drink enough fluids, these factors may also negatively affect a person's obesity level. By observing these factors, we may be able to adjust our models to more accurately portray the efficacy of each variable.

THANK YOU

Q&A