What is the **foundational** research in obesity and ML? In obesity and health informatics?

What is the most **recent** research in obesity and ML? In obesity and health informatics?

The critique we got from our proposal presentation was that this research doesn’t contribute something new enough. How do we address this?

How do we distribute the research tasks?

> Roy is familiar with SVM and KNN and Random Forest

> Mistiha is familiar with CNN

> Raunak is good with data visualization tools such as powerbi that can be used for reporting and predictive analysis. Also worked with python libraries such as Scikit-learn, pandas, numPy etc.

How do we distribute the writing of the paper?

First of all, we need to make a small alteration in the topic name, we remove the term “classifying” since we not classifying but making prediction of obesity.

Title: Classifying and Predicting Obesity Classes

Other potential titles…

* Classification of obesity classes

Literature Review

* We could divide by topic
  + ML and obesity
  + Health informatics and obesity
  + ML and health informatics
  + …
* Read 6 to 10 papers each
  + **Write a brief summary of research, result, and future direction [Feb 28]**
* Each of us could write ½ page about designated research topic

Abstract

Introduction   
Methods

Results

Discussion

Conclusion

**For Monday Feb 28th at 3:00 pm**

* **Have Aninda send dataset to everyone**
* **At least 6 paper and descriptions** 
  + **Written portion should include purpose of research, results of the research, and future direction from that research**
* **Exploratory analysis on dataset using the language of your liking** 
  + **How do variables affect each other**

Exploratory Data Analysis:

1. Bar diagram for Gender specific eating habits, physical condition, Obesity level.
2. Bar Diagram for Age specific eating habits, physical condition, Obesity level.
3. Correlation plot

Statistical Analysis;

1. Chi-square association between obesity with family history with overweight, FAVC,CAEC,CALC,SMOKE
2. Stepwise logistic regression for initial model selection
3. Wald test for variable selection
4. Goodness of fit test (Likelihood ratio test or Hosmer-Lemeshow test)

Classification Model:

1. Logistic regression
2. Support Vector Machine
3. K- Nearest Neighbor
4. Random Forest
5. Decision Tree
6. Naive Bayes
7. Neural Network

Here we have problems with these variables (FCVC,NCP,CH20,FAF,TUE) in our dataset. They are categorical but after the first 500 observations they have continuous observations which need to be fixed.