The dataset that we are going to work on is from Kaggle. It contains data on the vitals collected from individuals and their classification as either obese or not obese. The original dataset itself has 4 categories of obesity, however, we only consider the binary classification problem of either predicting an individual is obese or not obese (which also includes overweight, normal weight and underweight categories from the original dataset).

The purpose of this project is to compute various metrics for 3 different models: Random Forest classifier, K nearest neighbors classifier and a deep learning model (LSTM). We calculate these metrics individually on each of a 10 fold split of the original data set as well as present the results as an average of these metrics.

First, we shall perform an exploratory analysis of the dataset. It is small data set with 1000 observations. It contains details on age, height, weight, gender, BMI and physical activity level of a person.

A screenshot of a data

Description automatically generated

A screenshot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generated

A diagram of a baby

Description automatically generated with medium confidence

Next we shall define the function that computes the metrics that we would like to compute for each model.

A screenshot of a computer program

Description automatically generated

We also show the function that we used to create a model for each of the three types we study in this project.

A screen shot of a computer code

Description automatically generated

Once we can create a model we need a function that can calibrate and calculate the metrics of a given split of test and train.

A screenshot of a computer program

Description automatically generated

Now we run the main goal of this project.

A screenshot of a computer program

Description automatically generated

We see the following outputs for each iteration.

A screenshot of a computer code

Description automatically generated

A screenshot of a computer code

Description automatically generated

A white sheet with numbers and letters

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer program

Description automatically generatedA screenshot of a computer

Description automatically generated

We also define a simple function to compute average results.

A screenshot of a computer code

Description automatically generated

Which gives us the following results.

A screenshot of a computer

Description automatically generated