Code	**NAIVE_BAYES.PY**
Author	Nathaniel Heatwole, PhD (heatwolen@gmail.com) (GitHub) (LinkedIn)
Summary	Uses a naïve Bayes classifier (Gaussian) to predict body mass index (BMI)[1] level (discrete) using weight, height, and age (both from scratch and using sklearn).
Methods/ process	Naïve Bayes: - Supervised learning method and probabilistic classifier that predicts membership across two or more groups (Y) using one or more predictors (X). - Based around Bayesian inference (which itself is grounded in Bayes' theorem). - For each covariate (X), assumes the probability of each class (Y) is proportional to the probability density function (PDF) of a normal (Gaussian) distribution.[2] - Parameters of these distributions are typically estimated using maximum likelihood.[3] For a normal distribution, these are given by the empirical mean and standard deviation in the training data (group-level, by covariate). - Assumes the covariates (X) act independently to influence group membership (Y) (i.e., multiplicative probabilities).[4] - Weights the various probabilities (PDFs) using the empirical distribution of group membership in the training data (shares of observations in each group).[5] - The predicted probabilities are then renormalized so they sum to one.[6] Steps: 1. Import and clean training data. 2. Generate group-level descriptive stats (shares, mean, sigma) and use them to compute normal distribution PDFs for each covariate-group combination. 3. For each group level, take the product of: the empirical group shares (in the training data), and all of the normal distribution PDFs (one for each covariate). 4. Renormalize so all probabilities sum to one (rowwise).
Training data	BMI data – empirical health-related data for 741 persons (from Kaggle).
Output	Summary: - Training data descriptive stats - BMI level distribution – empirical, from scratch, using sklearn
Result	The results from scratch and using sklearn align perfectly. The overall accuracy rate of the model is about 90%.

[1] BMI is a height-normalized measure of weight. It is defined as: weight / height^2.

[2] This is analogous to the *likelihood* in Bayesian inference.

[3] Informally, this means choosing the model parameters so as to maximize the likelihood that the particular input data would be observed.

[4] Hence why the method is termed "naïve," as it does not allow for variable dependences or interactions.

[5] This is analogous to the *priors* in Bayesian inference.

[6] The normalizing factor is analogous to the *evidence* in Bayesian inference.