The task is to develop a health categorization framework that uses daily vital statistics and lifestyle patterns to identify patients as Good, Moderate or Poor health status. The original dataset contained data challenges such as heart rate and temperature values presented in pipe-separated format and blood pressure readings comprising both systolic and diastolic measurements. The extraction of valuable features required statistical computations to determine the mean values and standard deviation as well as minimum and maximum points from time-series information. Data became clearer through separation of blood pressure measurements into their systolic and diastolic components followed by averaging each segment. I converted the sleep quality categorical features to machine-readable format through one-hot encoding. The processed data included engineered features which converted unorganized fields into structured information before modeling could begin.

Before processing data, I checked for null entries then learned the dataset included no missing values in any of the columns. I implemented a backup strategy that set zero values in place of missing data to protect against future potential disruptions of the pipeline, yet this step was performed because no missing data existed in the current state of the information. The fallback method preserves data consistency and prevents data loss particularly during practical use when missing entries frequently occur. Next, I utilized StandardScaler to normalize all numeric features so the model would analyze inputs with equal scales and reduce the effect of features with wider data ranges. A special labeling function for health status classification was created through medical expertise-based threshold criteria. The classification of "Good" health applied when patients maintained vital signs inside their general health requirements (heart rate between 60–100 bpm alongside seven hours of sleep and 7000 steps). The labeling system assigned a "Poor" evaluation to health conditions where blood pressure amounts exceeded optimal ranges or physical activity remained minimal. The health status system labeled people's health condition as "Moderate" when their measurements existed between the two endpoint categories.

I selected the Random Forest Classifier to model this class problem because it works well with diverse data sets and generates information about feature importance. The model functions through a decision tree followed by result aggregation to enhance generalization levels and improve forecasting precision while reducing overfitting. I split my data into 80% training data and 20% testing data before training the model to generate subsequent assessment using classification report along with confusion matrix and feature importance plots. The metrics revealed model performance regarding its ability to segregate health categories because heart rate together with sleep duration and systolic blood pressure acted as primary influential factors. The task proved my competency to process sophisticated health data while presenting a learning method with clear procedures specifically designed for healthcare knowledge acquisition.