**Executive Report**

Max Jefferson, Jasmine Williams, Lahari Yelamanchili, & Jaelyn Young

Osteoporosis Risks & Prediction

**Overview**

Our streamlit app provides insights into the risks of osteoporosis based on both lifestyle and demographic factors. The primary objective of our application is to empower users, including healthcare professionals, researchers, and individuals concerned about osteoporosis, to gain valuable insights into the factors influencing osteoporosis prevalence. The model features include a dashboard overview, model description, user interaction form, data visualization, and performance metrics.

**Model Summary**

The dataset used to develop this model originally had 1,958 observations of 16 different variables. After removing an identifier variable, 15 of them remained. First, the dataset was split into 20% testing data and 80% training/validation data. The training/validation data was then split into 70% training data and 30% validation data.

Three initial models were created: a Logistic Regression Model (LR), a Random Forest Classification Model (RF), and a Decision Tree Classification Model (DT). The preliminary accuracy scores are as follows: Accuracy (LR): 0.5073059360730593, Accuracy (RF): 0.4999916977999169, Accuracy (DT): 0.5026857617268575. These accuracy scores are relatively low. A Grid Search method was then used to find the optimum hyperparameters for the Random Forest Classification Model (RF) and the Decision Tree Classification Model (DT); there were no hyperparameters to tune for the Logistic Regression Model (LR). From here, new models were created using the hyperparameters obtained from the Grid Search.

The accuracy scores of the new models are as follows: Accuracy (RF2): 0.8567413864674138, Accuracy (DT2): 0.8886550435865503. The F-1 scores are as follows: F-1 Score (RF2): 0.8378918844953693, F-1 Score (DT2): 0.877713768115942. Because the new Decision Tree Model has both the highest accuracy and F1 scores, this was the model chosen for the pipeline. In other words, the Decision Tree Model had the highest performance, so it was chosen to be *the model* for this application.

**Conclusions**

By visualizing data and model performance metrics, users can better understand the disease dynamics and make informed decisions regarding prevention, diagnosis, and treatment strategies.