### --------------------------------------------------------------------------------------------------------

# Random Forest - Spine Surgery

### --------------------------------------------------------------------------------------------------------

### Random Forest:

- an \*\*algorithm\*\*:

- based on \*\*Supervised Learning\*\*

- to solve \*\*Regression and Classification\*\*

- where machines \*\*learn with supervision\*\*

- \*\*input data are labeled\*\* and \*\*expected output data is known\*\*

- with key \*\*objective\*\*:

- \*\*to predict (classify) a categorical (qualitative) dependent output value (y)\*\* based on \*\*independent input variable(s) (x)\*\* -> for \*\*Classification problems\*\*

- \*\*to predict a continuous (quantitative) numeric dependent output value (y)\*\* based on \*\*independent input variable(s) (x)\*\* -> for \*\*Regression problems\*\*

- by \*\*choosing the feature that best split the data\*\* we try:

- \*\*to predict the class that is the mode of the classes of the individual trees\*\*

- \*\*to predict mean prediction of the individual trees\*\*

- \*\*Regression and Classification\*\* are a \*\*predictive modeling techniques\*\*

- A way \*\*to improve the predictive performance of a Decision Trees\*\*:

- \*\*constructs a collection (an ensable) of Decision Trees\*\* using bootstrap samples of the training dataset -> sampling from the training set with replacement

- training algorithm applies the general technique of bootstrap aggregating, or bagging, to tree learners

- Behave with \*\*“if this, then that” condition\*\*

- Can be \*\*used for larger datasets\*\*

- \*\*Hard to visualize the forest\*\* (a lot of trees)

### --------------------------------------------------------------------------------------------------------

### Project Objective: Predicting if spine surgery was successful

Create a model that allows to put in a few features of children who have had corrective spine surgery due to Kyphosis and returns back a prediction (classification) if the corrective spine surgery was successful. Information about the children who have had corrective spine surgery is in the dataset 'Kyphosis\_Data.csv'. The Kyphosis dataset has 81 rows and 4 columns.

The Kyphosis dataset contains the following columns:

- \*\*Kyphosis\*\* - present or absent after operation

- \*\*Age\*\* - age of children at time of operation (in months)

- \*\*Number\*\* - number of vertebrae involved in the operation

- \*\*Start\*\* - number of first (top-most) vertebrae that was operated on

\*\*Source\*\*: John M. Chambers and Trevor J. Hastie eds. (1992) Statistical Models in S, Wadsworth and Brooks/Cole, Pacific Grove, CA.

### --------------------------------------------------------------------------------------------------------

### Table of Contents:

1. File Descriptions

2. Technologies Used

3. Structure of Notebook

4. Executive Summary

#### 1. File Descriptions

- Random Forest - Spine\_Surgery.ipynb

- Kyphosis\_Data.csv

- README.md

#### 2. Technologies Used

- Python

- Pandas

- Numpy

- Matplotlib

- Seaborn

- Scikit-Learn

#### 3. Structure of Notebook

1. Import the Libraries

2. Load the Data

3. Exploratory Data Analysis

- 3.1 Check out the Data

- 3.2 Data Visualization

4. Data Preprocessing and Feature Engineering

- 4.1 Identify the variables

- 4.2 Dealing with Missing values

- 4.3 Dealing with the Non-numerical features

5. Train and Test the Random Forest Classifier model

- 5.1 Split the columns

- 5.2 Split the data into Training dataset and Testing dataset

- 5.3 Create the Random Forest Classifier model

- 5.4 Train / fit Random Forest Classifier model

- 5.5 Predictions from the model on Testing data

- 5.6 Evaluate the model on Testing data

- 5.6.1 Classification report

- 5.6.2 Confusion matrix

- 5.7 Feature importance

- 5.8 GridSearchCV

- 5.8.1 Create the Grid of parameters

- 5.8.2 Create the GridSearchCV model (Re-create the Random Forest Classifier model)

- 5.8.3 Train / fit the GridSearchCV model (Re-train / Re-fit the Random Forest Classifier model)

- 5.8.4 Predictions from the GridSearchCV model (Re-predictions from the Random Forest Classifier model) on Testing data

- 5.8.5 Evaluate the GridSearchCV model (Re-evaluate the Random Forest Classifier model) on Testing data

- 1. Classification report

- 2. Confusion matrix

#### 4. Executive Summary

TBA