

# Springboard Data Science Course

## Data Science Capstone Project 1

### Orthopedic Biomechanical Features

Michelle Ide - 12/29/2020

#### ~~~ MACHINE LEARNING ~~~

Data from Kaggle datasets were cleansed resulting in a single outlier removed. Exploration of data and hypothesis testing using  $\alpha = 0.5$ , confirmed Abnormal and Normal classifications of the data were statistically significant, the hypothesis test failed. Below summarizes the approach used for building machine learning models to predict classification.

#### Approach

The project uses 6 quantitative Features and 1 binomial Target to perform a classification of Normal/Abnormal using supervised learning techniques.

5 discriminative classification models were tested:

- ☐ Logistic Regression
- ☐ KNearest Neighbors
- ☐ Random Forest
- ☐ SVM
- ☐ Decision Tree

2 generative models tested:

- ☐ Naive Bayes
- ☐ LinearDiscriminantAnalysis

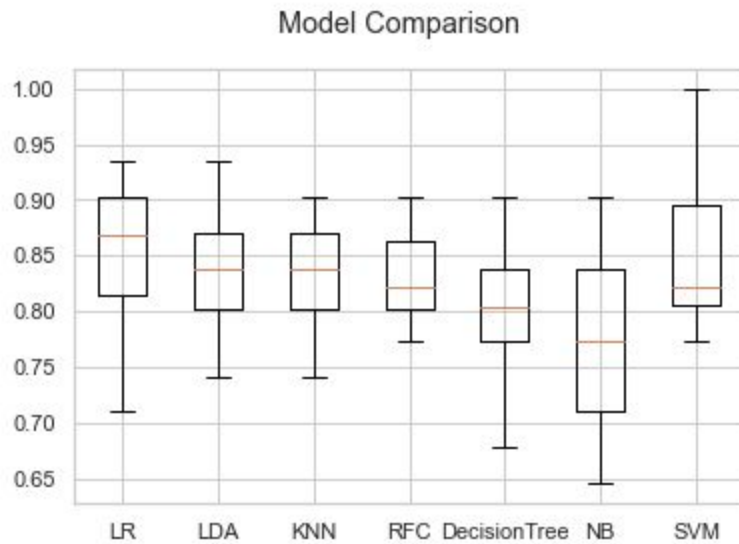
To prevent overfitting cross-validation was used to measure accuracy of each model.

#### Results

The initial measurements, prior to parameter tuning was performed with a KFold split of 10, measured on `corss_val_score` ( accuracy, SD):

- LogisticRegression: 0.851183 (0.073741)
- KNearestNeighbors: 0.834839 (0.053295)

- RandomForestClassifier: 0.828387 (0.041331)
- SVM: 0.850968 (0.065314)
- DecisionTree: 0.799355 (0.065759)
- NaiveBayes: 0.776452 (0.083024)
- LinearDiscriminantAnalysis: 0.838065 (0.058097)



Based on accuracy scores, further parameter tuning and testing was conducted on Logistic Regression and SVM.

Due to the low variance of RandomForestClassifier, additional parameter turning was performed and tested.