Yazan Khalaf

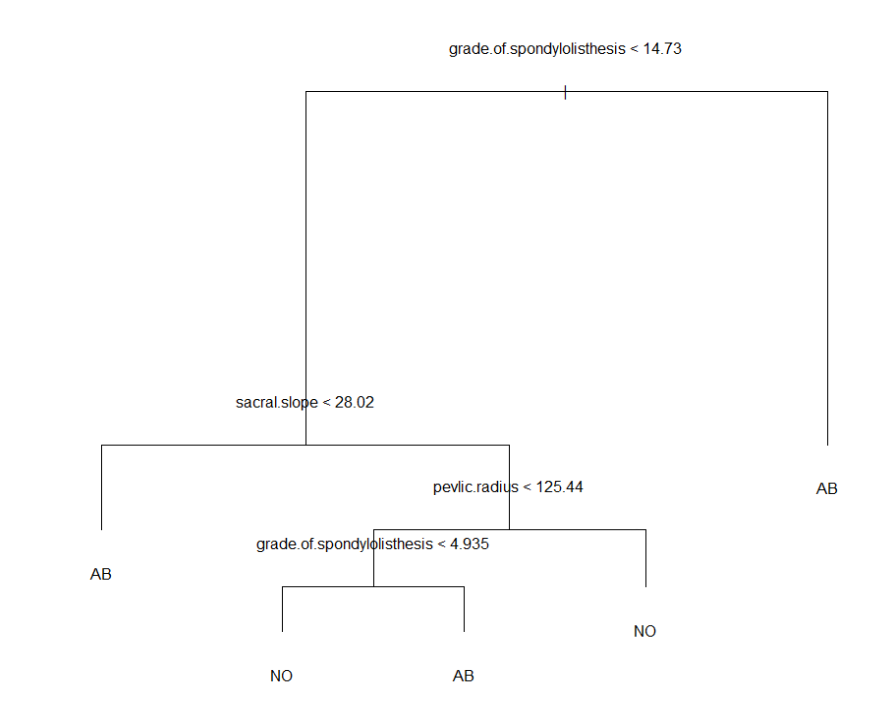
**Vertebral Column R Analysis**

What Are We Finding?

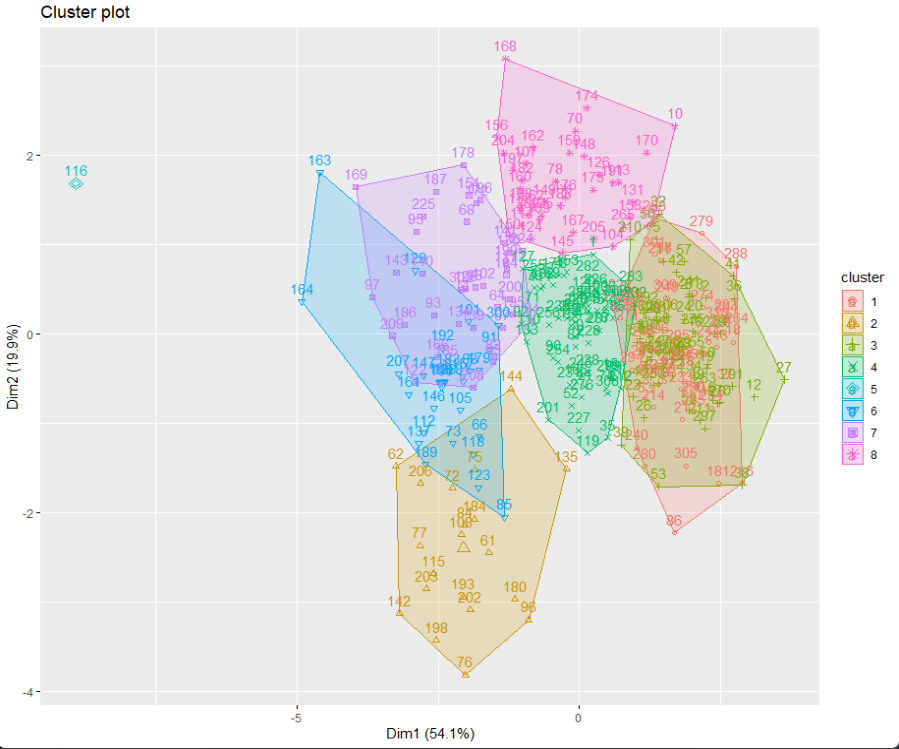
In this analysis I’ll be applying data mining techniques in R to data relating to lower back pain. Specifically, it contains observations of patients and their vertebral measurements, a total of 6 attributes were recorded. Each patient was also classified as normal or abnormal. Abnormal contains patients with a disk hernia or varying spondylolisthesis. I’ve chosen to apply classification and cluster analysis on my data. Classification can be very helpful here, in order to narrow down which measurements are most indicative of a lower back abnormality. I’m going to apply clustering in order to investigate whether the algorithm will cluster the data into their correct injuries.

What Steps Were Taken?

Initially I had to investigate the data. No null values were present and all the numerical values were appearing as such. Other than transforming the class variable into a factor, nothing more was needed prior to classification. I decided to utilize a decision tree and randomforest techniques, in order to create a simple and effective visual guide while also exploring the possibility of higher accuracy. For the best results I utilized pruning.

For clustering I immediately identified that scaling was necessary, especially considering the vast differences in ranges for the numerical attributes. I decided to utilize k-means clustering for its ability to set a specific set of clusters, in order to investigate whether the findings would reveal a similar result to the initial 3 back pain groups.

How Can We Spot Back Pain?

The decision tree narrowed down the important attributes from 6 to 3, those being spondylolisthesis, sacral scope, and pelvic radius and had an accuracy of 73%.

K-means allowed me to identify that the optimal amount of clusters was 8. This is insightful, considering the patients in the abnormality group originally split into only 2 types of injuries. This suggests that there might be more similarities between some groups of patients that we should consider important and further look into for treatment.