**Project proposal for ELE: 581**

Name: Muhammad Enayetur Rahman

Student ID: 100635221

Data set: [Vertibral Column Data Set](http://archive.ics.uci.edu/ml/datasets/Vertebral+Column)

Total 7 attributes are:

1. Pelvic Incidence 2. Pelvic Tilt 3.Lumber Lordosis Angle 4.Sacral Slope

5. Pelvic Radius 6. Grade of spondylolisthesis (SL) 7.Decision

Here is the basic statistical summary for each attribute. Attribute ‘Decision’ would be my dependent attribute and it has the Binary labels: Abnormal (AB) and Normal (NO).

|  |
| --- |
| **Pelvic.Incidence** |
| Min. : 26.15 |
| 1st Qu.: 46.43 |
| Median : 58.69 |
| Mean : 60.50 |
| 3rd Qu.: 72.88 |
| Max. : 129.83 |

|  |
| --- |
| **Pelvic.Tilt** |
| Min. :-6.55 |
| 1st Qu.:10.67 |
| Median :16.36 |
| Mean :17.54 |
| 3rd Qu.:22.12 |
| Max. :49.43 |

|  |
| --- |
| **Lumbar.Lordosis.Angle** |
| Min. : 14.00 |
| 1st Qu.: 37.00 |
| Median : 49.56 |
| Mean : 51.93 |
| 3rd Qu.: 63.00 |
| Max. : 125.74 |

|  |
| --- |
| **Sacral.Slope** |
| Min. : 13.37 |
| 1st Qu.: 33.35 |
| Median : 42.41 |
| Mean : 42.95 |
| 3rd Qu.: 52.69 |
| Max. : 121.43 |

|  |
| --- |
| **Pelvic.Radius** |
| Min. : 70.08 |
| 1st Qu.:110.71 |
| Median :118.27 |
| Mean :117.92 |
| 3rd Qu.:125.47 |
| Max. :163.07 |

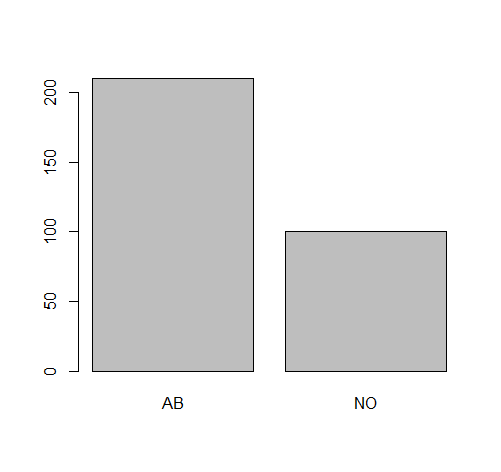
|  |
| --- |
| **Grade.of.SL** |
| Min. :-11.06 |
| 1st Qu.:1.60 |
| Median :11.77 |
| Mean :26.30 |
| 3rd Qu.:41.28 |
| Max. :418.54 |

|  |
| --- |
| **Decision** |
| AB:210 |
| NO:100 |

**Graphs of each independent variable:**

|  |  |
| --- | --- |
| **Attribute: Pelvic Incidence:** | **Pelvic Tilt** |
| **Lumbar Lordosis Angle** | **Sacral Slope** |
| **Pelvic Radius** | **‘Grade of SL’ Attribute:** |

Histogram of Dependent variable: **‘Decision’**



**Why using this data set?**

While Model building in WEKA, it is used Classify screen with Sequential Minimal Optimization (SMO) and found the summary below:

|  |  |  |  |
| --- | --- | --- | --- |
| Correctly Classified Instances | | 306 | 98.7097 % |
| Incorrectly Classified Instances | | 4 | 1.2903 % |
| Kappa statistic | 0.9705 | | |
| Mean absolute error | 0.0129 | | |
| Root mean squared error | 0.1136 | | |
| Relative absolute error | 2.9497 % | | |
| Root relative squared error | 24.2996 % | | |
| Total Number of Instances | 310 | | |

Also, when constructing Support Vector Machine model of dependent variable ‘**Decision’** with *linear Kernel* in R studio it is found that the values predicted and the actual values for the ‘Decision’ attribute almost matched.