**Data Mining Project #3**

***Biomechanical Features of Orthopedic Patients***

*By*

*Jigar Parmar*

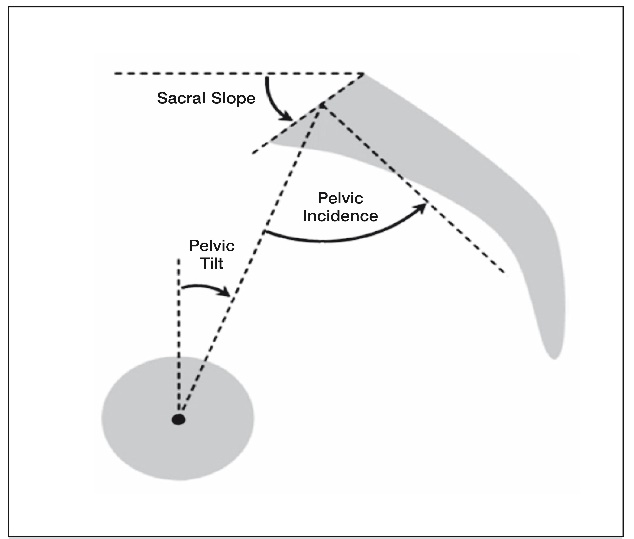
*Harish Parasuraman*

*Vaishnavi Krishna Guda*

Let us discuss the variables in detail:

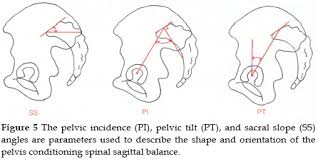
**Pelvic Incidence:**

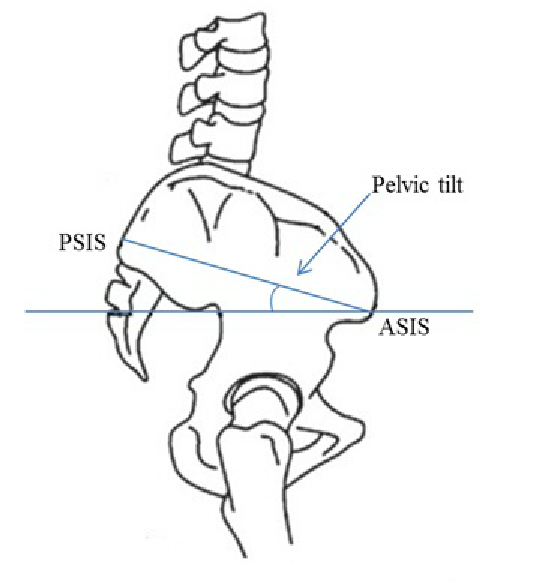
Advancement of the human clade to acquire an erect stance and bipedal movement has been made conceivable through a progression of skeletal morphological changes. The pelvic vertebra has played a critical part in this change by experiencing dynamic extending and retroversion. Upkeep of an upright pose requires the body's focal point of gravity to fall over a limited zone between the feet while keeping up a flat look with minimal energy expenditure. The pelvic incidence (PI) is measured as an angle formed by two vectors. The pelvic incidence is a fixed morphological parameter, whose value remains constant throughout adult life. It is an algebraic sum of two dynamic angles: the pelvic tilt and the sacral slope. (Ramchandran S, n.d.).



**Pelvic Tilt:**

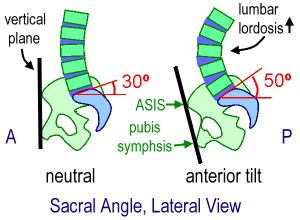
Pelvic tilt is an anteroposterior motion of the pelvis around an imaginary axis in the frontal plane. The most common tendency for pelvic rotation in the seated posture is the posterior pelvic tilt. Anterior pelvic tilt is a type of postural distortion that occurs when the pelvic bones are rotated forward and down as if it was a bucket of water, pouring water onto your feet. (NMT, 2008)





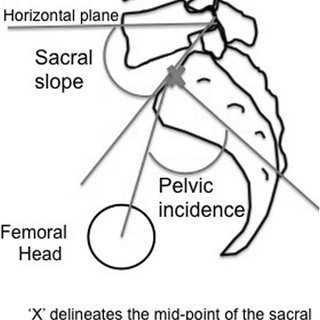
**Lumbar Lordosis angle:**

Lumbar lordosis has for some time been contemplated, and its curvature is related with different elements, for example, thoracic curvature, age, gender, pelvic curve, among others. Studies have been led meaning to gauge lumbar and spinal sections curvatures. The interest of vertebral bodies and inter-vertebral plates on lumbar lordosis arrangement has not been considered. A supine lateral lumbar spine radiograph is a very accurate means of measuring lordotic angles. The LSA is one of such angles, and can be used in the investigation, treatment and follow-up of low back disorders. (DAMASCENO, n.d.)



**Sacral Slope:**

The sacral slope(SS) measure the Sacro-pelvic orientation in the sagittal plane, being evidenced in the lumbosacral lateral view. SS is defined as the angle between the upper horizontal plateau and S1, while Pelvic Tilt is the angle between the line connecting the midpoint of the upper plateau of S1 and the center of femoral rotation with a vertical line. The degree of the sacral slope determines the position of the lumbar spine, since the sacral plateau forms the base of the spine. Pelvic Tilt and Sacral Slope, on the other hand, are dependent positions, because they depend on the angular position of the sacrum/pelvis in relation to the femoral head, which changes Vertical reference line (VRL) .PT/SS ratio is also affected by the bending and lumbosacral-pelvic extension. (Roussouly P, 2014)



**Pelvic Radius:**

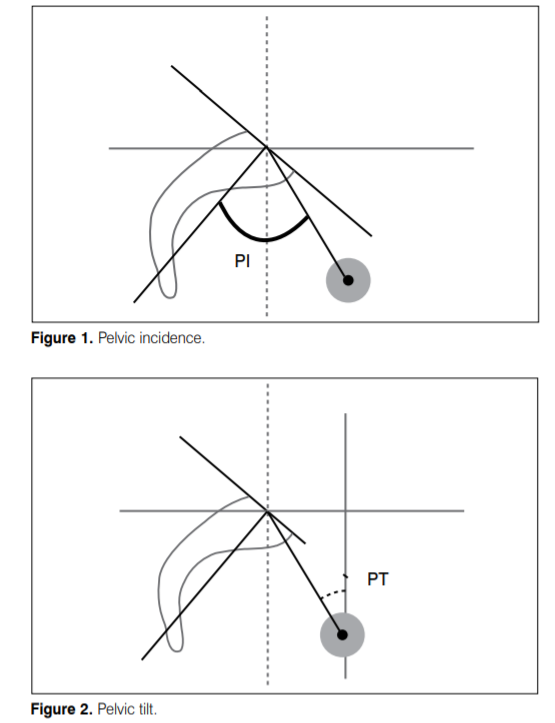
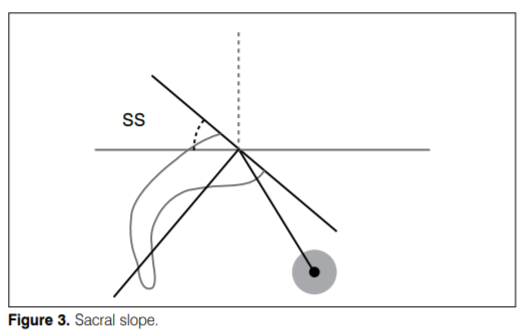
Pelvic radius technique involves measurements based on a line drawn between the hip axis and the posterior corner. The angle formed between this line and the sacral endplate, PR, is a developmental measure of Sacro-pelvic morphology. (•, n.d.)

**Grade of spondylolisthesis**:

An acquired anterior displacement of one vertebra over the subjacent vertebra, associated with degenerative changes, without an associated disruption or defect in the vertebral ring is known as Spondylolisthesis.



Difference between PI, SS and PT:



# References

•, J. C. (n.d.).

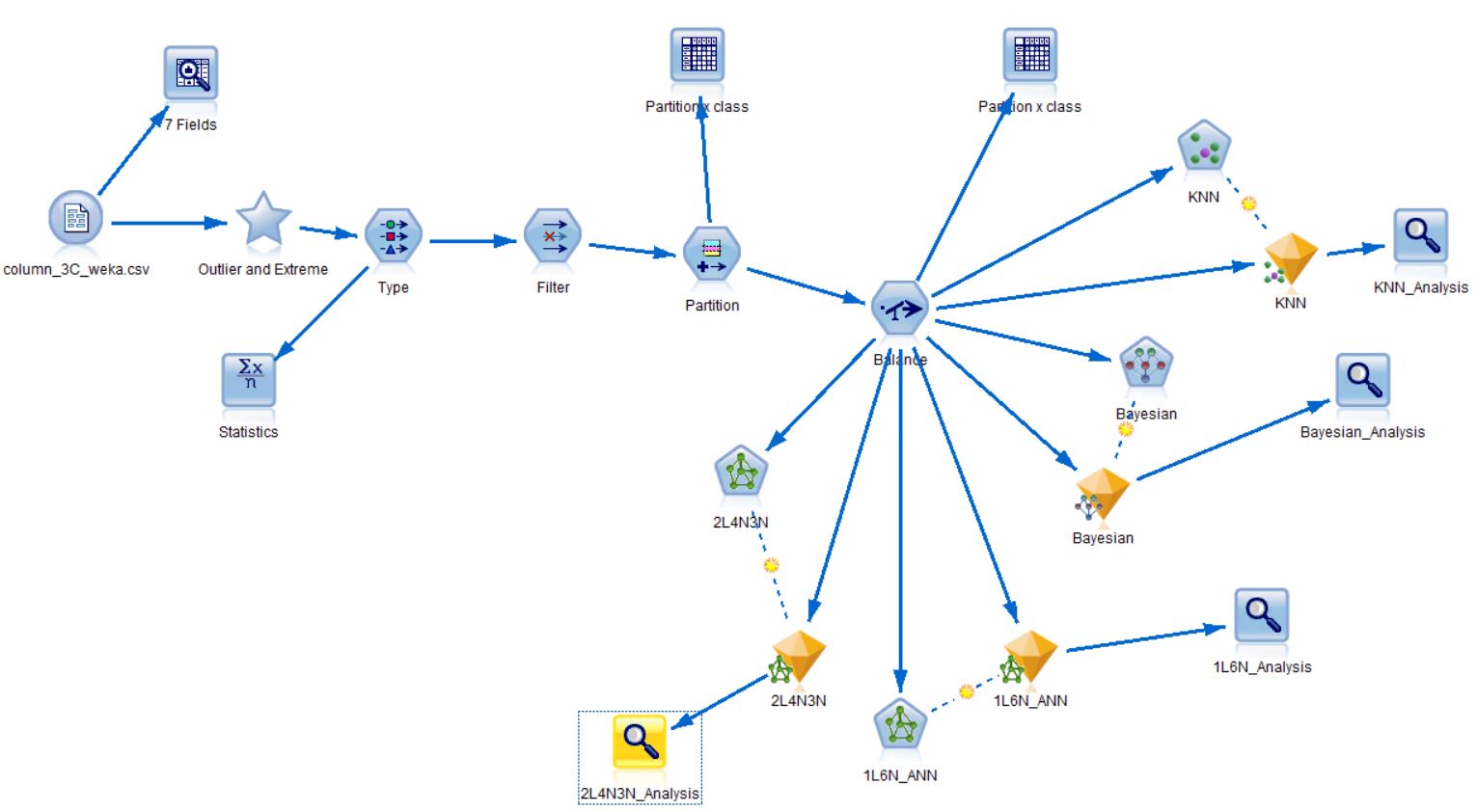
DAMASCENO, L. H. (n.d.).

NMT, S. V. (2008). Retrieved from www

Ramchandran S, B. A. (n.d.). *imedpub Journal*.

Roussouly P, G. S. (2014). *r ev b ras or top .* . Retrieved from www

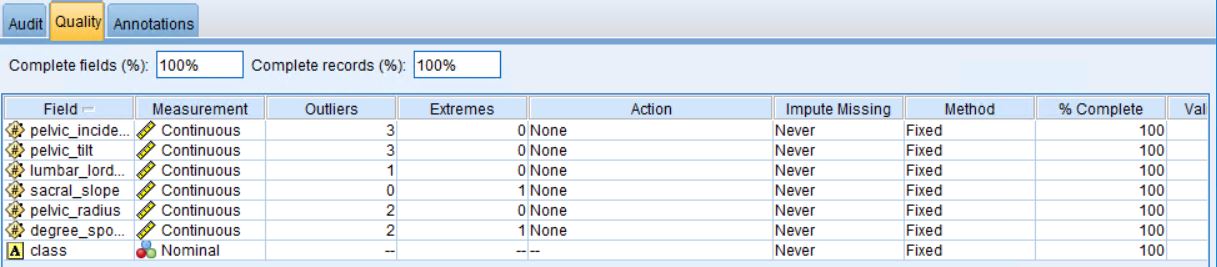
SPSS Modeler Implementation:



**Fig 1**

We are going to implement the data set using SPSS Modeler to predict the condition of a patient i.e. normal, dischernia, Spondylolisthesis as in the above figure **Fig 1**. Since the given data set is in the form of .csv, so we will use the .var file available in the source field of the SPSS modeler to load the data set. Before performing any kind of analysis, we checked for the missing values, outliers and correlations among the independent variables.

Below figure displays the data audit node where we can see that however there are no missing values in the data, but there are outliers and extremes present in the data set.

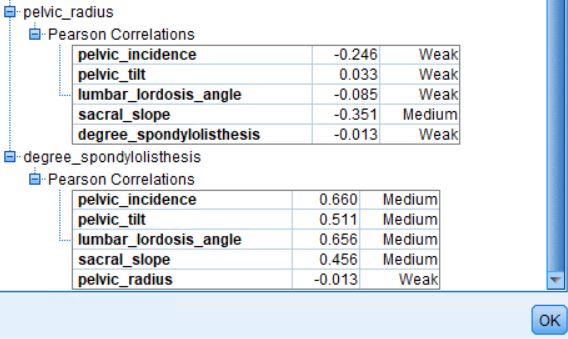
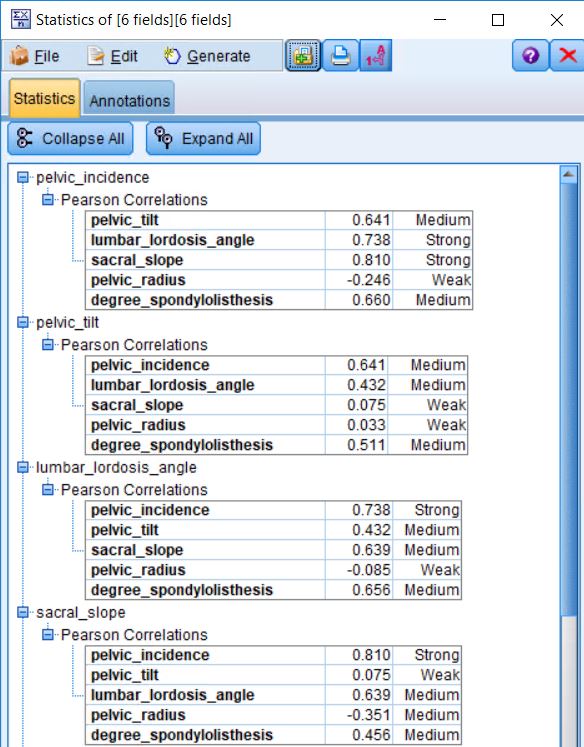


**Fig 1.A**

In order to remove the outliers and extremes, we used Scores in the “Action” column. This forces the outliers and extremes to become a normalized data.

We generated a Super Node for outliers and extremes which can be displayed in the main figure **Fig 1** thereby passing our data set through this super node.

Now, the correlation among the variables are checked as in the below figure **1.B**

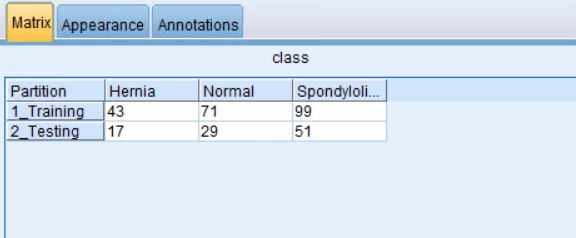


**Fig 1.B**

From the above figure, we can see that there exists strong correlation between few independent variables which cannot be allowed during the classification process of the data. In order to fix this, since there are only 6 variables are present, from the **fig 1.B**, we can see that the strong correlation exist between “pelvic\_incidence” and “lumbar\_lordosis\_angle” and between “pelvic\_incidence” and “sacral\_slope”.

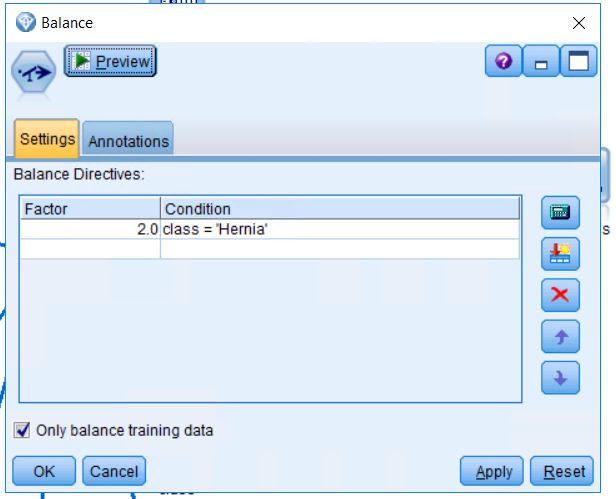
For now, we have removed “pelvic\_incidence” as an input to the model. So now, this fixes the correlation issue. Now the data is partitioned 70% - 30%, training and testing respectively.

The imbalance data, if any, among the dataset, is checked. So, in order to perform that, we attached a matrix node which is available in the output field of SPSS modeler to the Partition node. The following figure shows the output of the first matrix node.



**Fig 1.C**

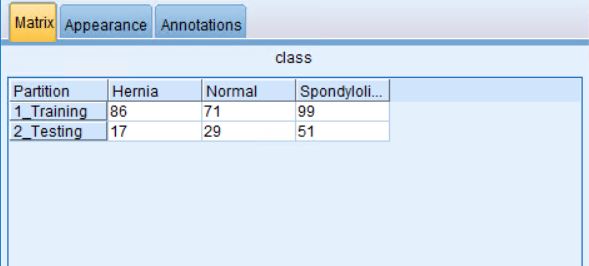
So from the above figure, we can see that number of patients with Spondylolisthesis is significantly more when compared to the number of patients with Hernia. (Approximately in the ration of 2:1). So again we cannot use the imbalanced data set for the classification process as it will generate unstable results. So in order to fix this, we have used balance node to balance out our data set. The configuration of balance node is as shown in the below figure:



**Fig 1.D**

Once the data is balanced, we again check for the partition to check if the data is still unbalanced or not.

For that, we again used the matrix node and below figure show the output of that.



**Fig 1.E**

From the output, we can see that the dataset has been balanced out.

Finally, our data set is clean enough to go through classification process.

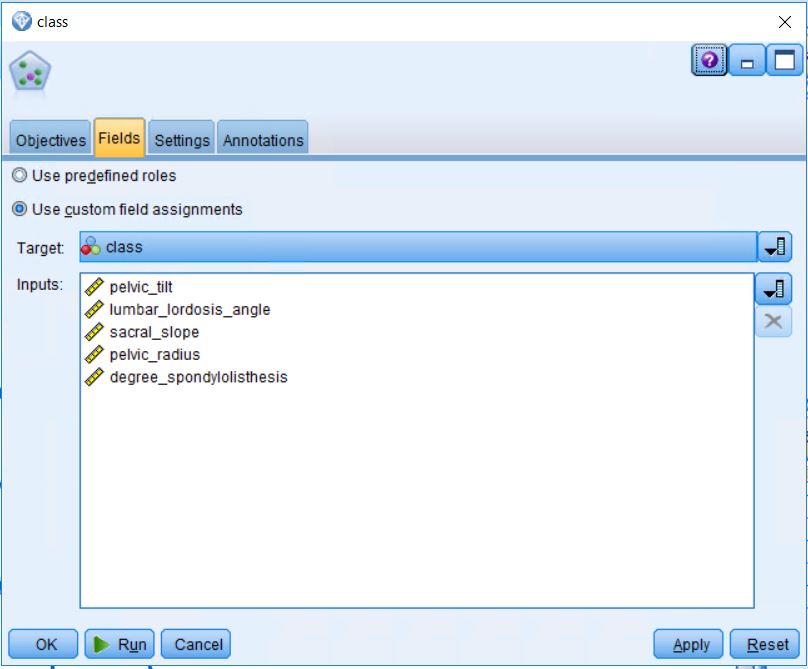
Now, as we have studied, Classification can be done in 3 ways:

1. Using K-Nearest Neighbor Analysis (KNN)
2. Bayesian Network
3. Artificial Neural Network (ANN)

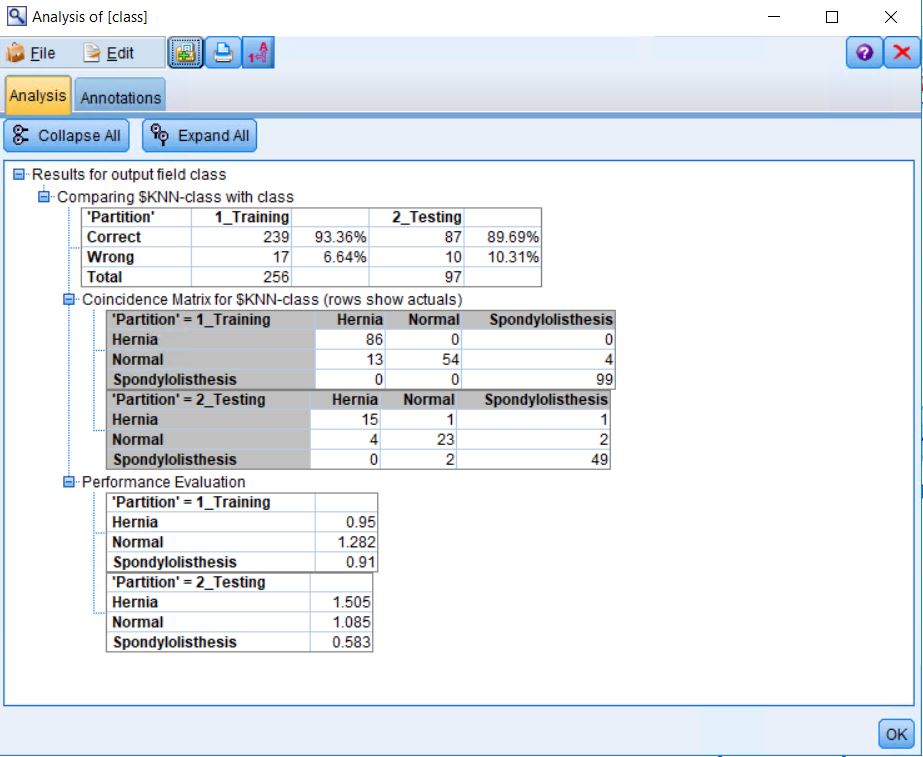
We have used all 3 techniques mentioned above for our classification process.

1. **K-Nearest Neighbor Classification (KNN):**

Since we have already partitioned our data into 70-30%, we have directly passed our data to KNN model. In our dataset the target is our class which contains 3 values Normal, DiscHernia and Spondylolisthesis and other variables are our input variables.



Following figure shows the output analysis of KNN model.



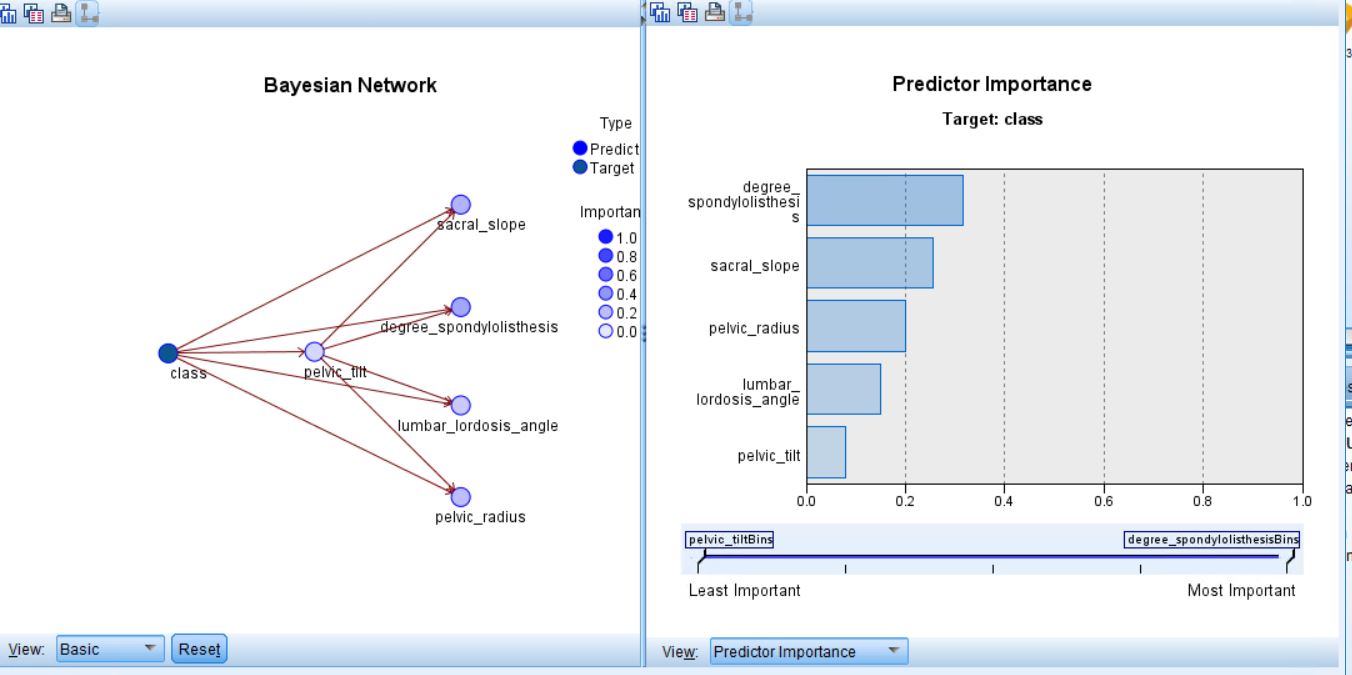
**Fig KNN\_1**

From the above analysis we can see that, this model is able to classify in a correct manner by 89.69%. This means that if this model will give 89 times correct classification out of 100 trials.

1. **Bayesian Network:**

The next model that we are going to use is Bayesian Network. For that, we need to pass the clean-partitioned-data through Bayesian model.

Again the target is class with all the other variables as input, the data being partitioned already into 70-30%. The following figures show the output of the Bayesian Network which was created on the basis of the given data set.



**Fig Bayesian\_1**

The above figure shows the Bayesian and the predictor importance for the classification process.

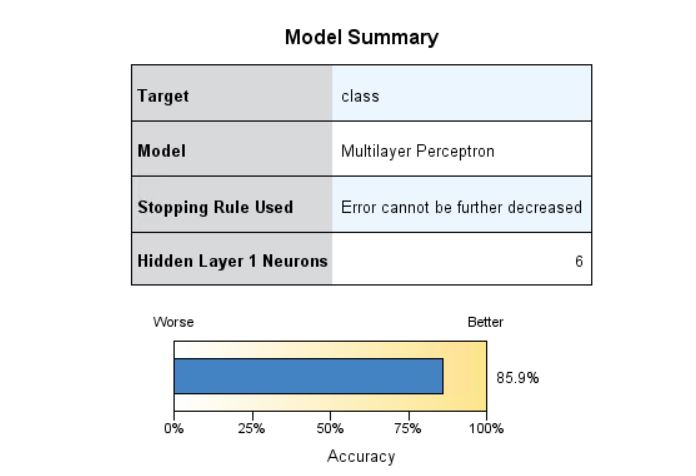
**Note:**

**We were trying to run the analysis node, but the output is not getting displayed. We tried multiple times but it not displaying.**

1. **Artificial Neural Network (ANN):**

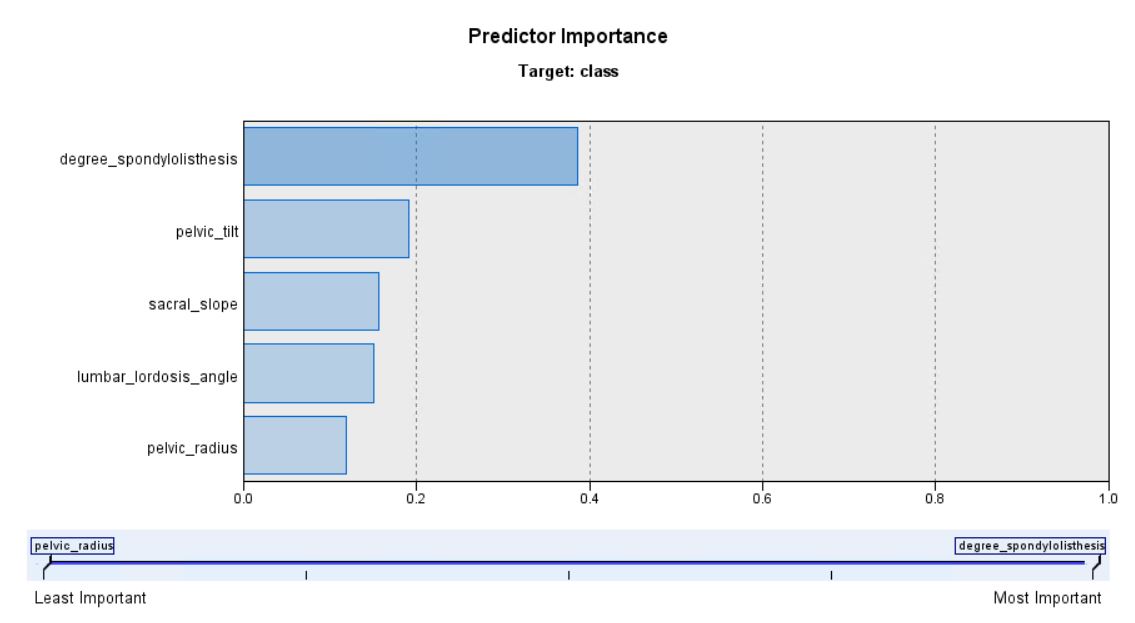
Our 3rd classification is through ANN. For this classification, we let the model decide for internal layers and neural. (Auto compute)

The output of ANN is as displayed below:



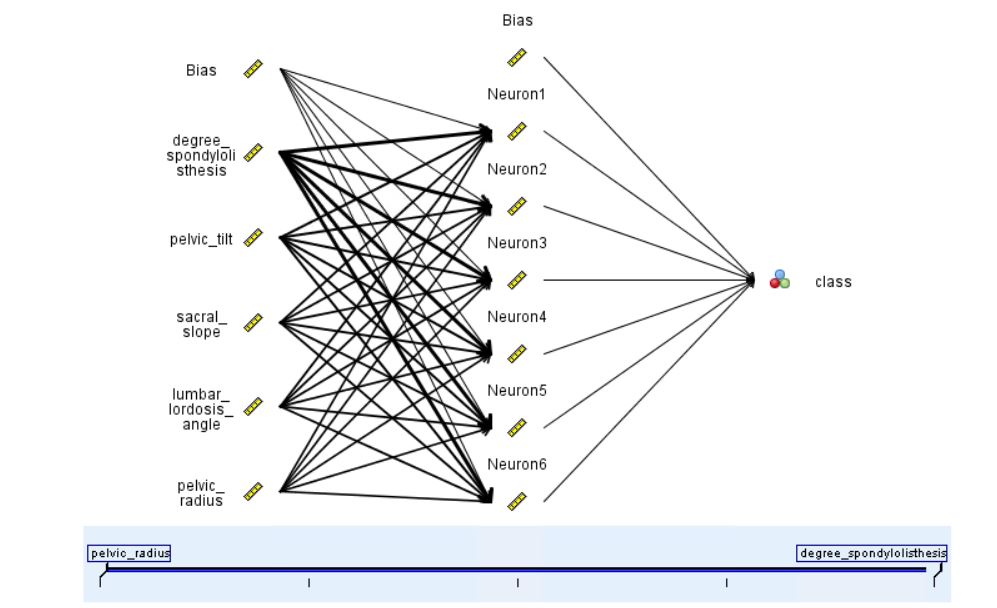
**Fig ANN\_1**

From the above figure, it is 85.9% accurate.



**Fig ANN\_2**

The above figure shows the Predictor importance

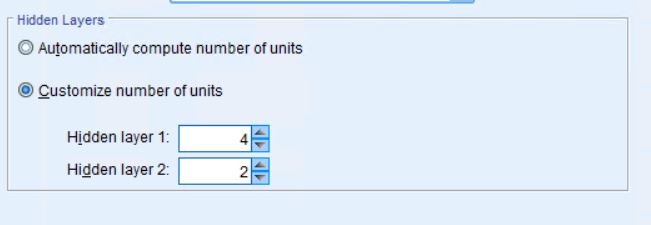


**Fig ANN\_3**

The above figure shows the Neural Network when the model decides the internal layers and neural.

Now again ANN is implemented, this time by giving the internal layers and neural by us.

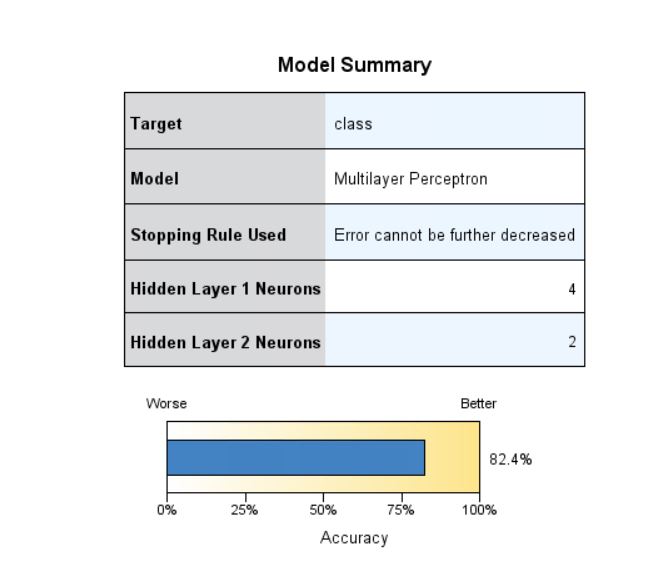
The below figure displays the layers given:



From the above figure, the hidden layers are given as 4 and 2 for layer 1 and layer 2 respectively.

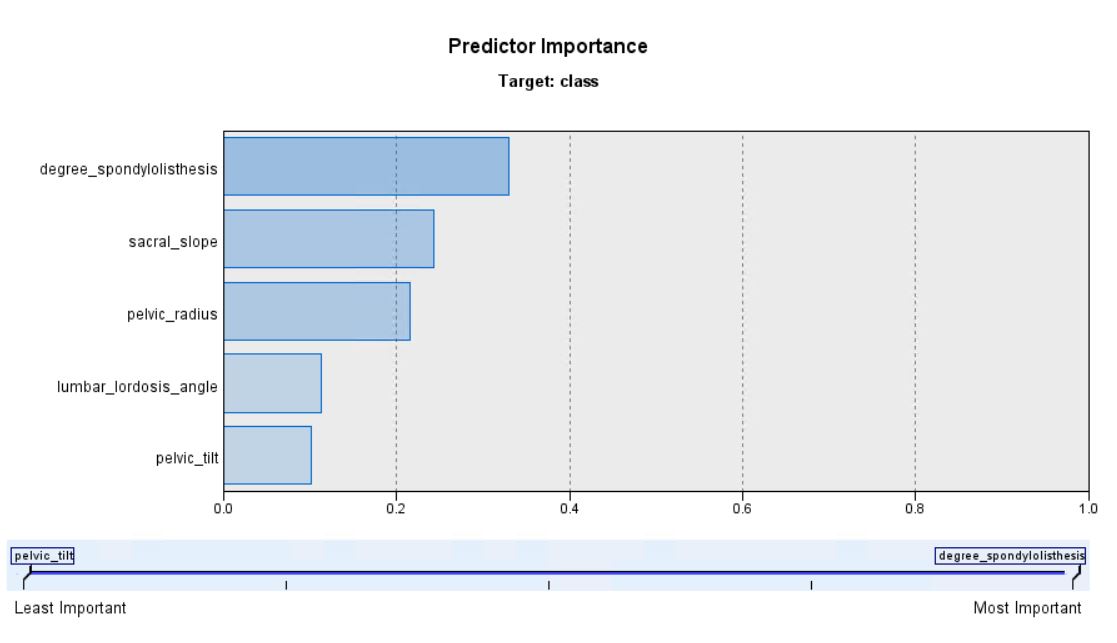
After, we, deciding the number of hidden layers, the ANN classification are implemented.

The below figures illustrates the classification:



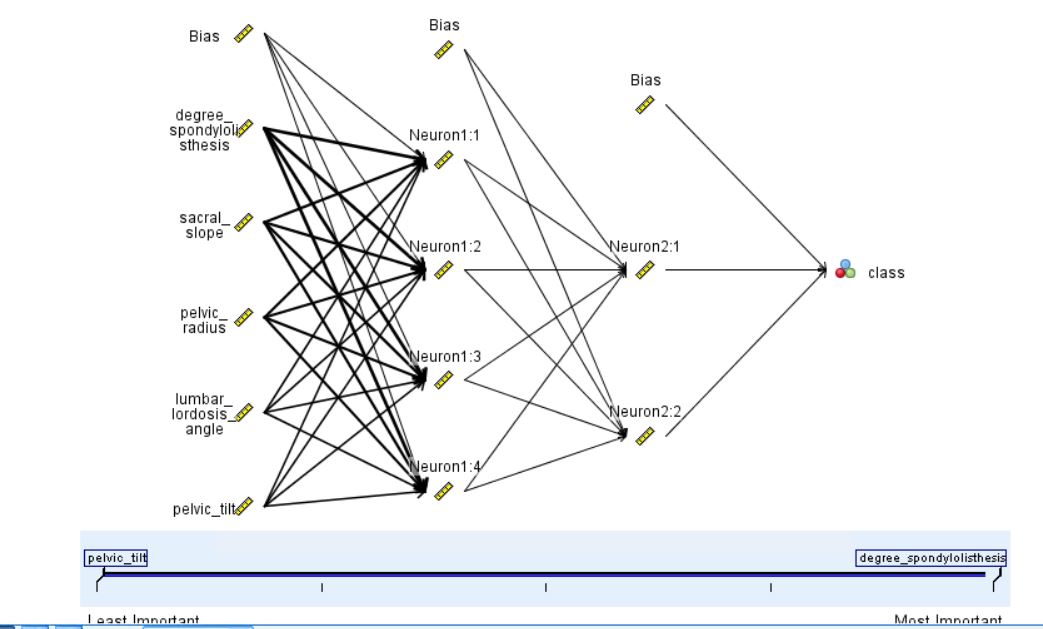
**Fig ANN\_4**

From the above figure, the accuracy is 82.4%



**Fig ANN\_5**

The above figure displays the Predictor Importance



**Fig ANN\_6**

The Bayesian Network when the number of hidden layers is given is displayed as above.