Springboard Capstone 1

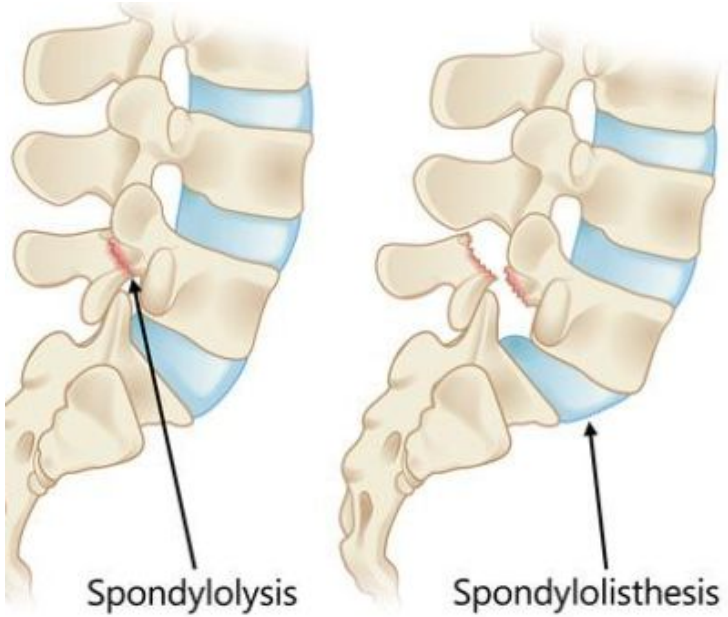
~~~Data Science Career Track~~~

Biomechanical Features of Orthopedic Patients

CLASSIFICATION
of
Spondylolisthesis
~~~

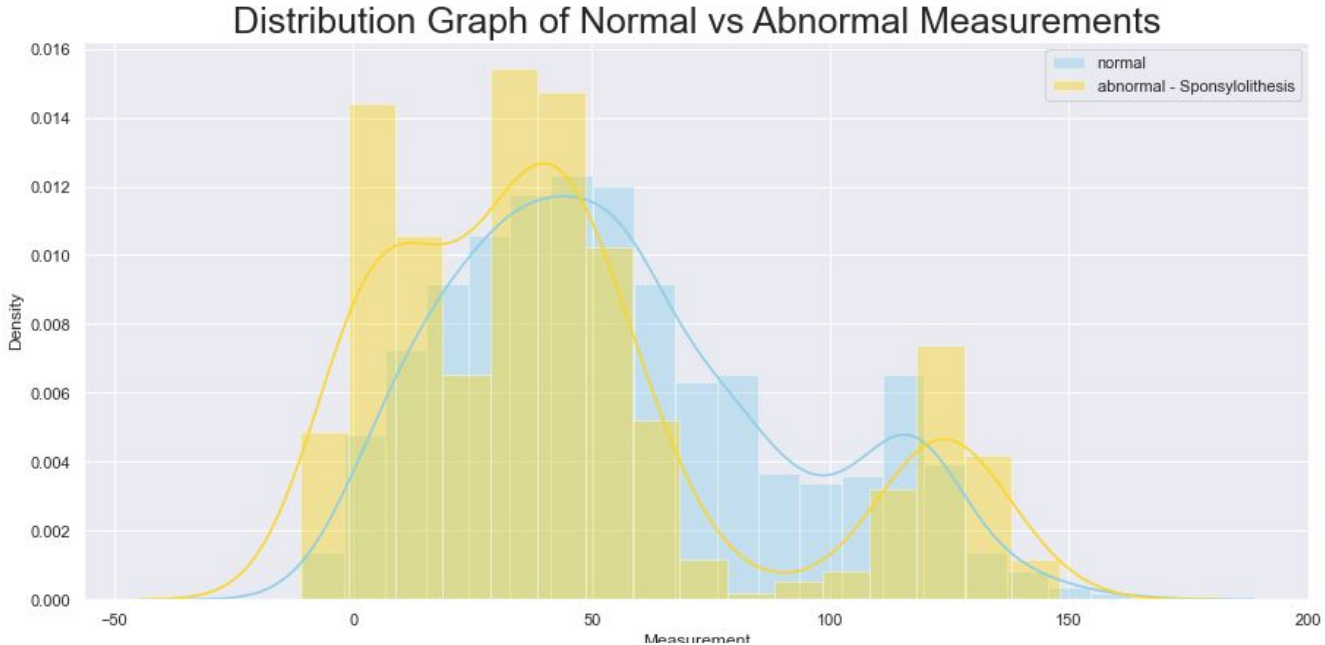
Painful lumbar (back) issues are primarily diagnosed on x-ray results

The problem is - Not all issues are as obvious as seen below



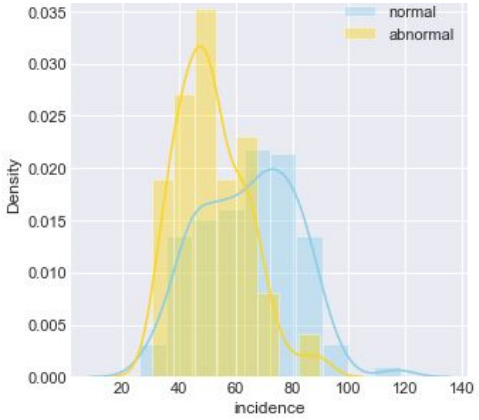
Reproduced from Cavalier R, Herman MJ, Cheung EV, Pizzutillo, PD: Spondylolysis and spondylolisthesis in children and adolescents: I. diagnosis, natural history, and nonsurgical management J Am Acad Orthop Surg 2006; 14: 417-424

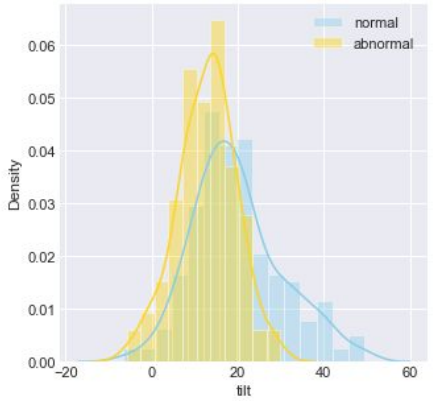
Abnormal measurements overlap Normal measurements.

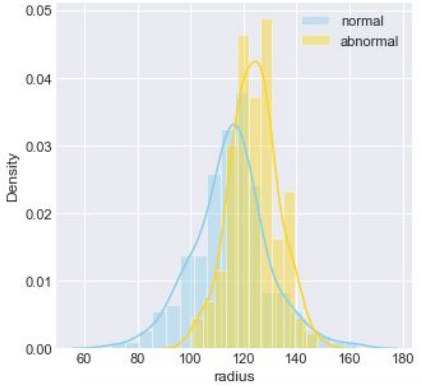


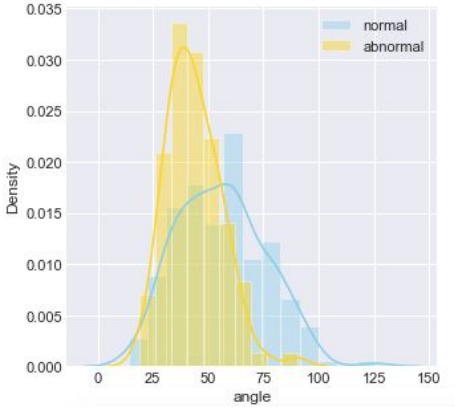
It appears the high-bills you've paid for radiology results may have been well earned.

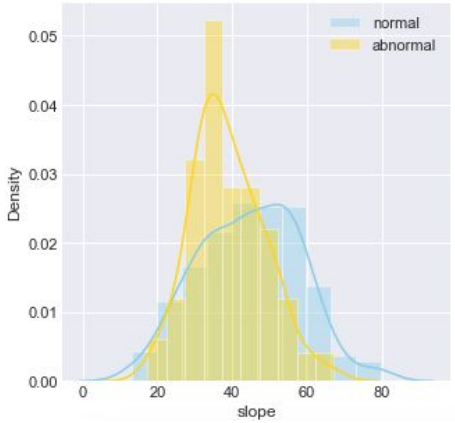
Proper diagnosis requires comparison of several features - *simultaneously*

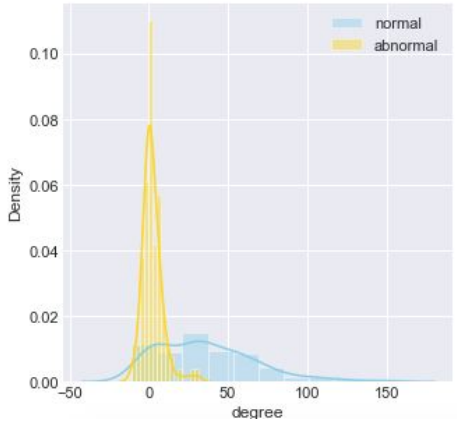












The physician bears the stress to simultaneously compare multiple measurements.

pelvic_incidence	pelvic_tilt numeric	lumbar_lordosis_angle	sacral_slope	pelvic_radius	degree_spondylolisthesis	class
63.027818	22.552586	39.609117	40.475232	98.672917	-0.254400	Abnormal
39.056951	10.060991	25.015378	28.995960	114.405425	4.564259	Abnormal
68.832021	22.218482	50.092194	46.613539	105.985135	-3.530317	Abnormal
69.297008	24.652878	44.311238	44.644130	101.868495	11.211523	Abnormal
49.712859	9.652075	28.317406	40.060784	108.168725	7.918501	Abnormal

Machines are perfectly suited to assist.

Project: Classification of Spondylolisthesis

Problem:

Diagnosis requires simultaneous comparison of multiple features.

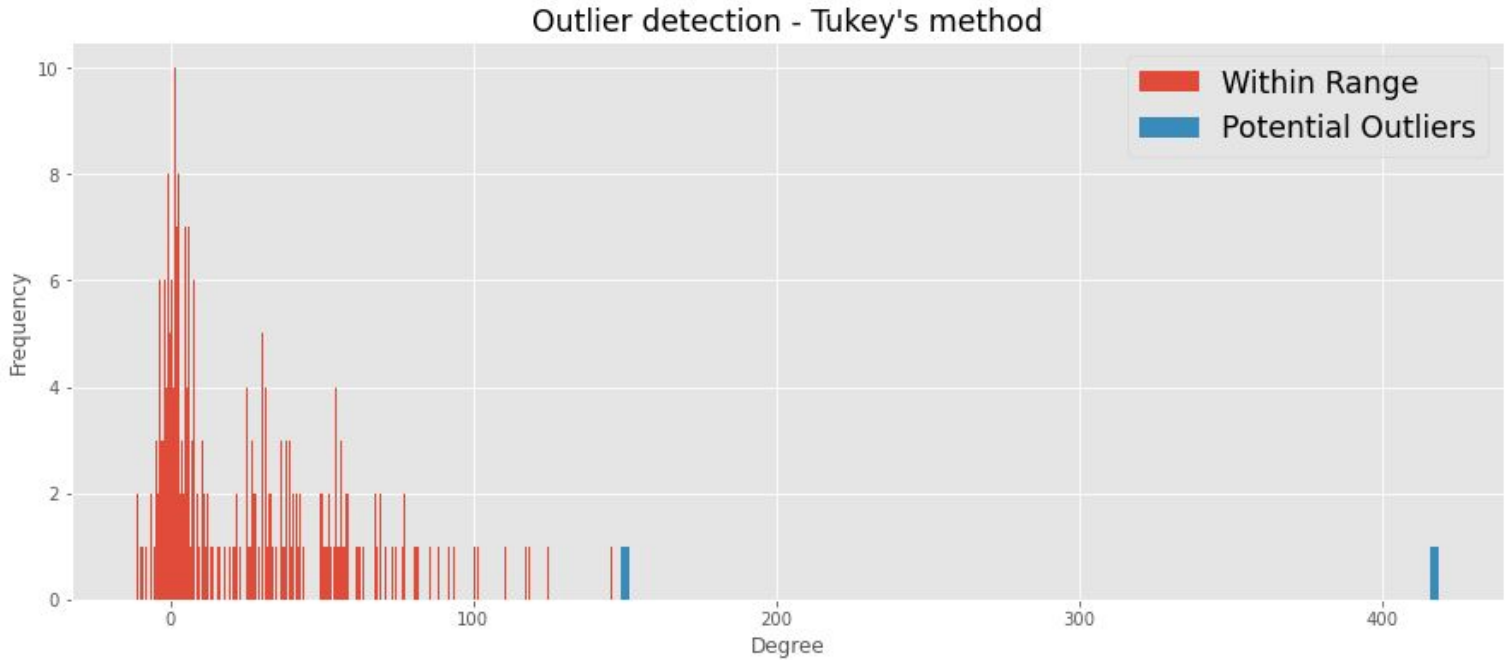
Solution:

Machine learning quickly classifies and sorts saving time and improving accuracy.

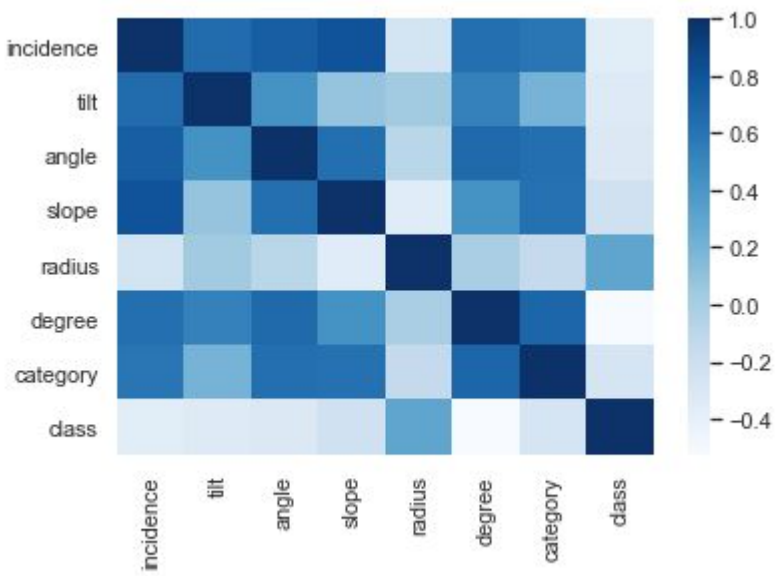
The Data

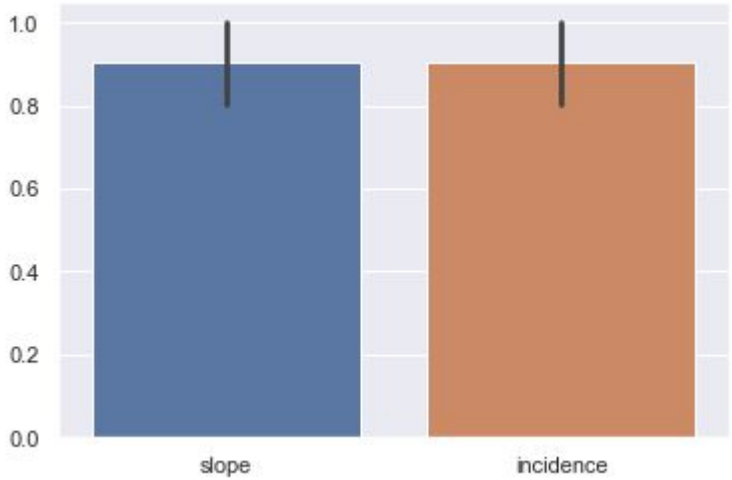
'Tukey's' algorithm was used to locate ***potential*** outliers.

1 outlier (410) was >2 times as large as the next data point (150) - it was removed.



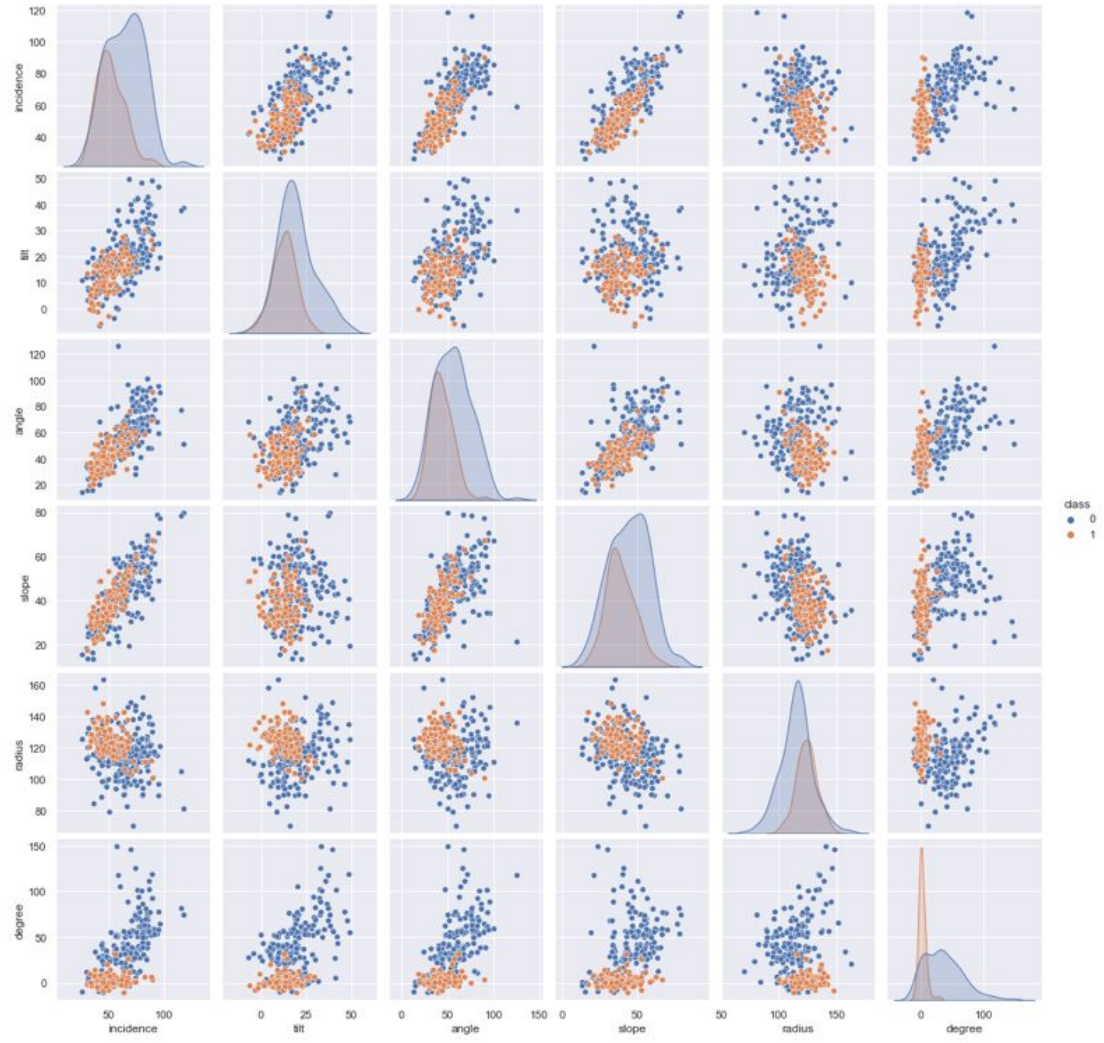
Correlations plotted to look for overly-correlated features





'slope' and 'incidence' were highly correlated, close to the 90% cut-off range.

A correlation module ran to remove overly-correlated features (>90%). All passed, none removed.



Correlations of features

Comparing the separation between "Normal" (orange) and "Abnormal" (blue).....

suggests that added weight might be useful on 'degree' and 'radius' to improve the model.

The Models

4 Models were used

- Logistic Regression
- Guassian Naive Bayes
- KNearest Neighbors
- Random Forest

Logistic Regression & The Base

A basic Logistic Regression was used to determine the base accuracy of 0.87

Additional tuning of C parameter using GridSearch did not change the results

Gaussian Naive Bayes

Due to the bi-model and right-skewed normal shape of some features - a Gaussian NB was tested with results less than the base

KN Neighbors

KN Neighbors returned the same accuracy after tuning the n_neighbors to =10

Tuning of parameters: algorithm, leaf_size, metric, metric_params did not improve accuracy.

Random Forest

Random Forest provided improved results of 88% once the n_estimator and random_state were tuned in tandom.

A module was created to loop through a range of values, by plotting the results we see the optimal values.