ORTHOPEDIC PATIENT CLASSIFICATION  
REPORT

PACE University

Seidenberg School Of Computer Science and Information Systems

CS627 – Artificial Intelligence

Spring 2025

Professor Dr. Soheyla Amirian

Student: Kien Chung

UID: U01956429

# INTRODUCTION

## OVERVIEW

A machine learning model was developed to predict orthopedic patient conditions based on six biomechanical features of the pelvis and lumbar spine. The dataset underwent preprocessing and multiple machine learning models were evaluated to ensure accurate classification between normal and abnormal spinal conditions. This approach aims to support data-driven diagnosis and improve clinical decision-making.

Biomedical dataset was built by Dr. Henrique da Mota during a medical residence period in the Group of Applied Research in Orthopaedics (GARO) of the Centre Médico-Chirurgical de Réadaptation des Massues, Lyon, France.

Kaggle Source:

<https://www.kaggle.com/datasets/uciml/biomechanical-features-of-orthopedic-patients>

Original Source – UC Irvine Machine Learning Repository:

<https://archive.ics.uci.edu/dataset/212/vertebral+column>

## OBJECTIVES

The data have been organized in two different but related classification tasks.

* The first task consists in classifying patients as belonging to one out of three categories: ‘Normal’, ‘Herniated Disc’, or ‘Spondylolisthesis’.
* For the second task, the categories Herniated Disc and Spondylolisthesis were merged into a single category labelled as 'Abnormal'. Thus, the second task consists in classifying patients as belonging to one out of two categories: ‘Normal’ or ‘Abnormal’.

## REQUIREMENTS

Python IDE

Python libraries & packages:

* pandas
* numpy
* matplotlib
* seaborn
* scikit-learn

# DATASETS

## DATAFRAMES

“**df\_2C.csv**”: has 2 classes ‘Normal’ and ‘Abnormal’.

“**df\_3C.csv**”: has 3 classes ‘Normal’, ‘Hernia’, and ‘Spondylolisthesis’.

## STRUCTURES

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **df\_2C.csv** | |  | **df\_3C.csv** | |
| X | Y |  | X | Y |
| Pelvic Incidence | Normal Abnormal |  | Pelvic Incidence | Normal  Hernia  Spondylolisthesis |
| Pelvic Tilt |  | Pelvic Tilt |
| Lumbar Lordosis Angle |  | Lumbar Lordosis Angle |
| Sacral Slope |  | Sacral Slope |
| Pelvic Radius |  | Pelvic Radius |
| Grade of Spondylolisthesis |  | Grade of Spondylolisthesis |

Notes: Six biomechanical features of both datasets have the same structure and values.

## DATA DISTRIBUTION

Total patients (310)

* Normal patients (100)
* Abnormal patients (210)
  + Patients with disc hernia (Herniated Disc) (60)
  + Patients with Spondylolisthesis (150)

Notes: The dataset contains approximately twice as many records for abnormal patients as for normal patients. Because of the class imbalance, the model probably performs better at predicting abnormal cases than normal ones.

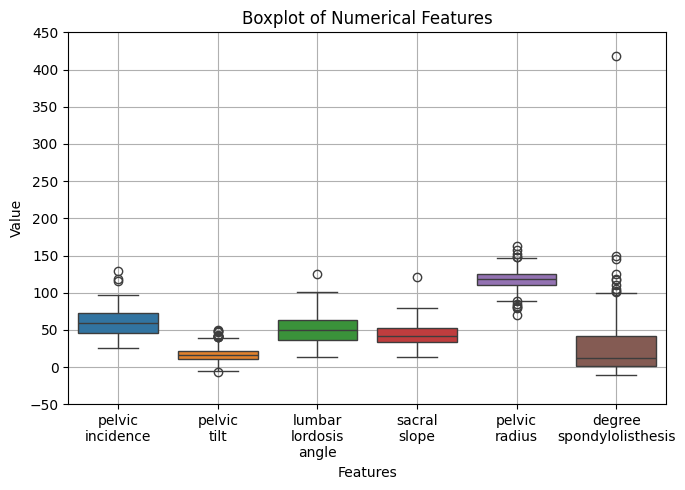
## DATA CONDITION

The dataset was examined for overall quality and found to be clean and well-structured. There are no missing or null values across any of the features, ensuring completeness of the data. Duplicate records were checked and none were found, confirming the uniqueness of each observation.

Additionally, the target variable (y) contains consistent and correctly labeled outcome classes, with no spelling variations or case inconsistencies. This clean state of the dataset provides a reliable foundation for subsequent modeling and analysis.

# EXPLORATORY DATA ANALYSIS & VISUALIZATION

## FEATURE DISTRIBUTION



The boxplot shows the distribution of six biomechanical features. Most features have a relatively symmetric spread with some outliers, especially in degree of spondylolisthesis, which exhibits extreme values and high variability. Pelvic radius has the tightest distribution, while pelvic tilt and sacral slope show moderate spread.

Overall, the features vary in scale and contain outliers that may need preprocessing before modeling.

# DATA PREPROCESSING

## FEATURE SCALING

The dataset contains six numerical biomechanical features. To prepare the data for modeling, the first four features were scaled using StandardScaler, which standardizes values based on the mean and standard deviation.

However, the last two features, ‘pelvic radius’ and ‘degree of spondylolisthesis’, exhibited significant outliers, as seen in the boxplot analysis. To address this, RobustScaler was applied to those features, as it is less sensitive to outliers and scales data based on the median and interquartile range.

This mixed-scaling approach ensures consistent feature scaling while minimizing the influence of extreme values.

## LABEL ENCODING

For the binary classification dataset df\_2C, the target variable consists of two classes: Normal (0) and Abnormal (1). One-hot encoding was applied to this target, converting it into a binary vector format suitable for certain algorithms that benefit from categorical expansion.

In contrast, the multiclass dataset df\_3C includes three classes: Normal (0), Hernia (1), and Spondylolisthesis (2). For this case, label encoding was used to retain the ordinal class representation, which is appropriate for multiclass classification tasks where the labels are already numeric and do not require further transformation.

# MACHINE LEARNING MODELING

## DATA SPLITTING

The dataset was split into training and testing sets using an 80/20 ratio with stratified sampling to preserve class distribution.

The training set contains 248 samples with 6 features, and the testing set includes 62 samples, ensuring a balanced representation of both classes in each subset using stratify.

## MODEL SELECTION & TRAINING

For this dataset, algorithms such as Random Forest, Decision Tree, and Naïve Bayes will be used to train models and compare their performance to identify the most effective model.

For tree-based models such as Random Forest and Decision Tree, feature scaling was not applied, as these algorithms are inherently insensitive to the scale of input features.

However, Naïve Bayes relies on probability calculations that assume normally distributed features, making feature scaling important for accurate results. Therefore, scaling was specifically applied for the Naïve Bayes model to ensure proper feature contribution and improved performance.

# RESULTS, COMPARISON, & INTEPRETATION

## DATASET df\_2C

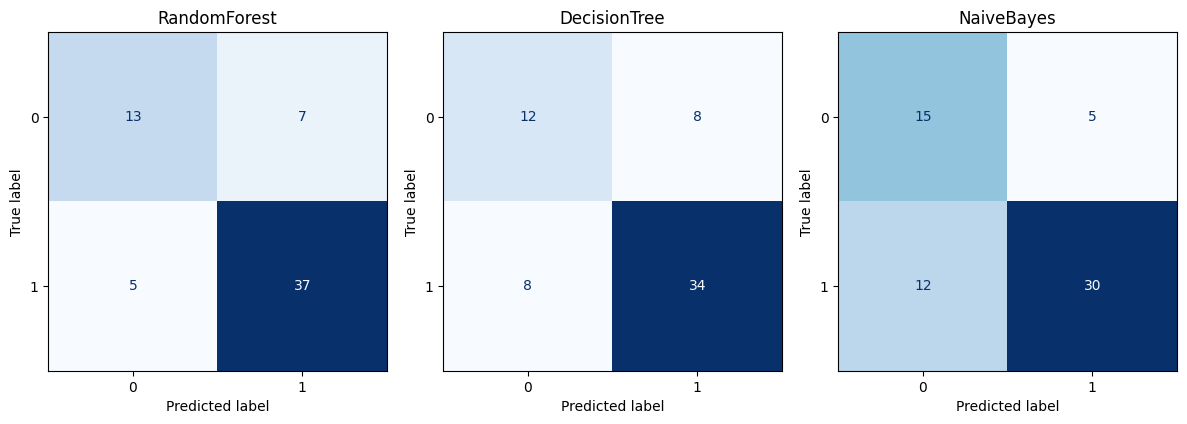
### SCORES

Random Forest - Accuracy: 0.8065, Precision: 0.8409, Recall: 0.8810, F1 Score: 0.8605

Decision Tree - Accuracy: 0.7419, Precision: 0.8095, Recall: 0.8095, F1 Score: 0.8095

Naïve Bayes - Accuracy: 0.7258, Precision: 0.8571, Recall: 0.7143, F1 Score: 0.7792

### CONFUSION MATRIX



Labels: Normal (0) and Abnormal (1).

## DATASET df\_3C

### SCORES

Random Forest - Accuracy: 0.8226, Precision: 0.8213, Recall: 0.8226, F1 Score: 0.8216

Decision Tree - Accuracy: 0.7258, Precision: 0.7609, Recall: 0.7258, F1 Score: 0.7391

Naive Bayes - Accuracy: 0.8871, Precision: 0.8836, Recall: 0.8871, F1 Score: 0.8844

### CONFUSION MATRIX

A diagram of a variety of blue squares

AI-generated content may be incorrect.

Labels: Normal (0), Hernia (1), and Spondylolisthesis (2).

## INTEPRETATION

Binary Classification (df\_2C — Normal vs Abnormal):

* Random Forest shows the best overall performance with high recall (88.10%), indicating it correctly identifies most abnormal patients. Its F1 score (86.05%) reflects a good balance between precision and recall. The confusion matrix confirms this, showing only 5 false negatives and 7 false positives, suggesting strong generalization.
* Decision Tree performed moderately well (F1 score: 80.95%) but had more misclassifications compared to Random Forest, including 8 false positives and 8 false negatives, implying less stability in separating the two classes.
* Naïve Bayes, despite having the highest precision (85.71%), had lower recall (71.43%), meaning it missed more abnormal cases (12 false negatives). This could be due to sensitivity to feature distributions, even after scaling.

Random Forest offers the most reliable predictions for abnormal cases, which is critical in medical contexts where missing an abnormal diagnosis is costly.

Multiclass Classification (df\_3C — Normal, Hernia, Spondylolisthesis):

* Naïve Bayes performed best, with an accuracy of 88.71% and F1 score of 88.44%. The confusion matrix shows near-perfect classification of spondylolisthesis (class 2) and very few misclassifications among the other two classes, suggesting that the model benefited from well-scaled features and distinct class boundaries.
* Random Forest also performed strongly (accuracy: 82.26%), especially in predicting spondylolisthesis, but showed some confusion between normal and hernia classes, as reflected by 5 hernia cases misclassified as normal.
* Decision Tree had the weakest performance (accuracy: 72.58%) and higher misclassification rates across all three classes, particularly mixing normal and hernia cases.

Naïve Bayes effectively captures differences between all three classes and handles the multiclass task better, likely due to the clean feature preprocessing and its probabilistic nature.

# CONCLUSION

For the binary (0/1) classification task (df\_2C), Random Forest outperformed other models with an accuracy of 80.65%, precision of 84.09%, recall of 88.10%, and an F1 score of 86.05%, making it the most effective at distinguishing between normal and abnormal cases.

In the multiclass classification task (df\_3C), Naïve Bayes achieved the best performance with an accuracy of 88.71%, precision of 88.36%, recall of 88.71%, and an F1 score of 88.44%, indicating strong and balanced classification across all three classes: normal, hernia, and spondylolisthesis.