Obstructive Sleep Apnea

- Case Study -

Jaime Pérez Sánchez

Problem Description

What challenge are we trying to solve?

Problem Description

Symptoms:

- Excessive daytime sleepiness
- Loud Snoring
- Morning headache
- Not rested after sleeping
- Abrupt awakenings
- High blood pressure
- ...

Problem Description

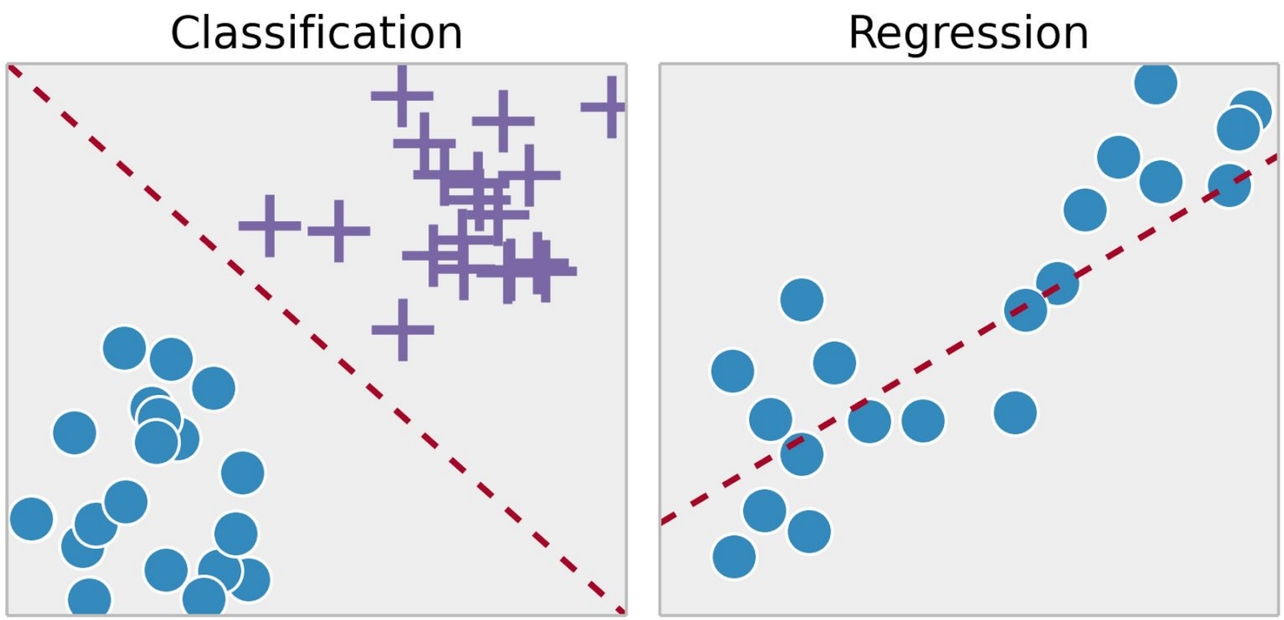
Risk Factors:

- Smoking
- Diabetes
- High blood pressure
- Neck circumference > 40 cm
- Be over 40 years of age
- Male
- ...

Machine Learning Approach

Objective: Apnea-Hypopnea Index (AHI)

Supervised Learning Problem:



Machine Learning Approach

Experimental Setup:

- Python (3.7) in an Anaconda's Virtual Environment
- Libraries:
    - Numpy
    - Pandas & Pandas-profiling
    - Scikit-Learn & XGBoost & CatBoost
    - Matplotlib & Seabon

Machine Learning Approach

Methodology:

1. Data Acquisition
2. Data Wrangling
3. Evaluation Metrics and Protocols
4. Model Selection and Training
5. Model Testing and Results
6. Hyperparameter tuning and Model Deployment

1.

Data Acquisition

Clinical dataset

2.

Data Wrangling

Describing and preparing the data to feed the Machine Learning models.

2. Data Wrangling

Pandas-Profiling: Report file

Selected columns:

- Patient (index)
- Gender (categorical)
- Weight (numerical)
- Height (numerical)
- Age (numerical)
- Smoker (categorical)
- Cervical Perimeter (numerical)
- BMI (numerical)
- AHI (target - numerical)

2. Data Wrangling

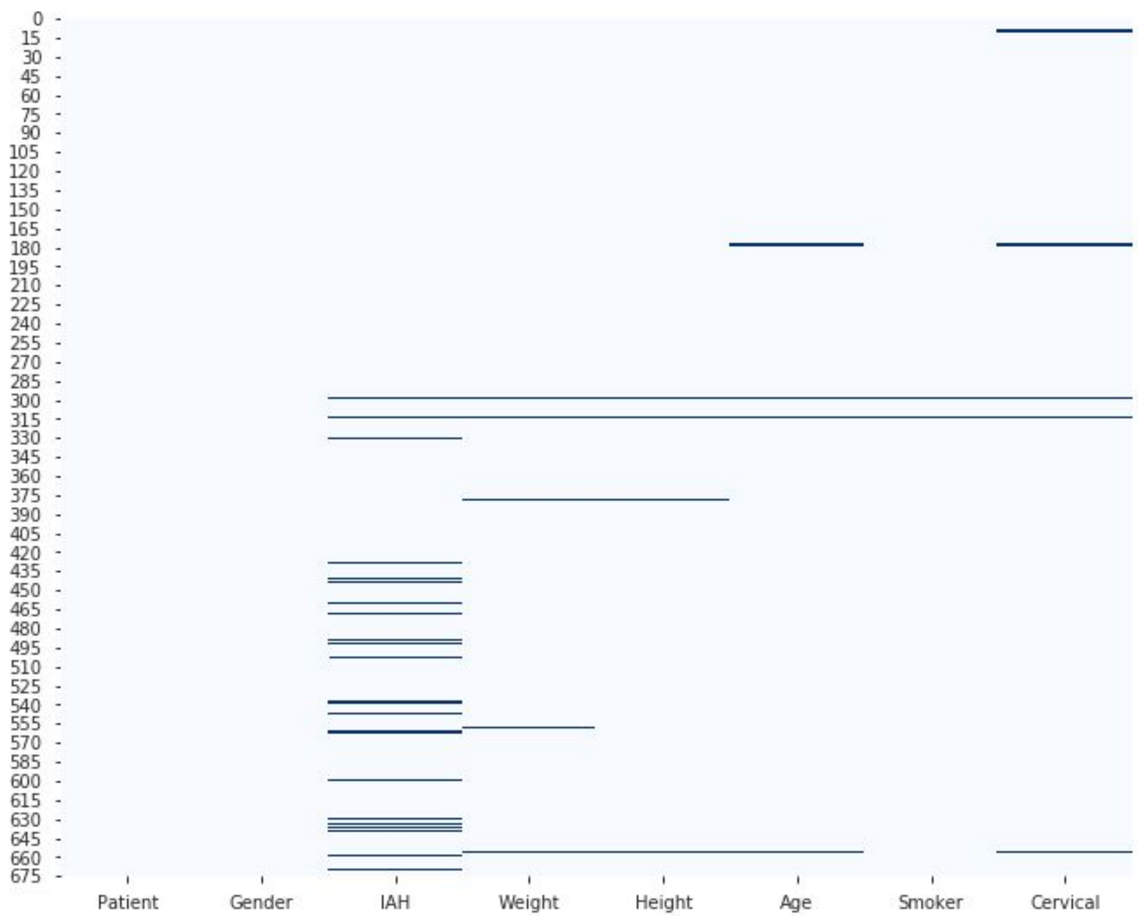
Missing Values:

- Replace 'ns' and -1 values with NumPy
- Drop NaN values with Pandas

	Patient	Gender	Weight	Height	Age	Smoker	Cervical	AHI
NaNs	0	0	7	6	5	3	5	34
-1	0	0	1	1	3	0	7	0

2. Data Wrangling

Missing Values:



2. Data Wrangling

Encoding Categorical Variables:

Label Encoding

Food Name	Categorical #	Calories
Apple	1	95
Chicken	2	231
Broccoli	3	50

→

One Hot Encoding

Apple	Chicken	Broccoli	Calories
1	0	0	95
0	1	0	231
0	0	1	50

2. Data Wrangling

Encoding Categorical Variables:

‘Smoker’

Label	Code
Non-smoker (“no”)	0
Former smoker (“antiguo”)	1
Light smoker (“poco”)	2
Smoker (“si”)	3

‘Gender’

...	Gender == Male	Gender == Female
...	1	0
...	0	1
...	0	1
...	1	0

2. Data Wrangling

Feature Engineering:

Creating new variables from available data

- $$BMI = \frac{Weight_{[kg]}}{Height^2_{[m]}}$$

- log(AHI+1)

2. Data Wrangling

‘OSA’ for Classification Models:

Label	Code
Healthy (AHI ≤ 10)	0
Severe (AHI ≥ 30)	1

3.

Evaluation Metrics and Protocols

Methods for evaluating the performance and generalization of the models

3. Evaluation Metrics and Protocols

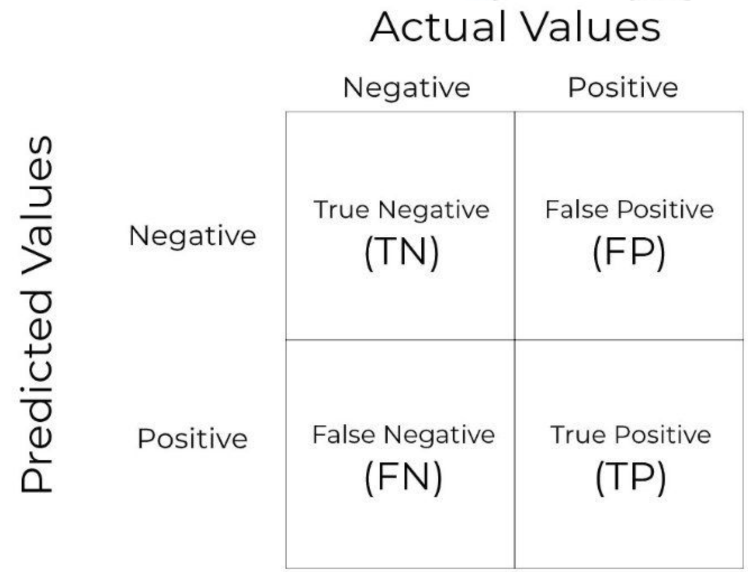
Regression metrics:

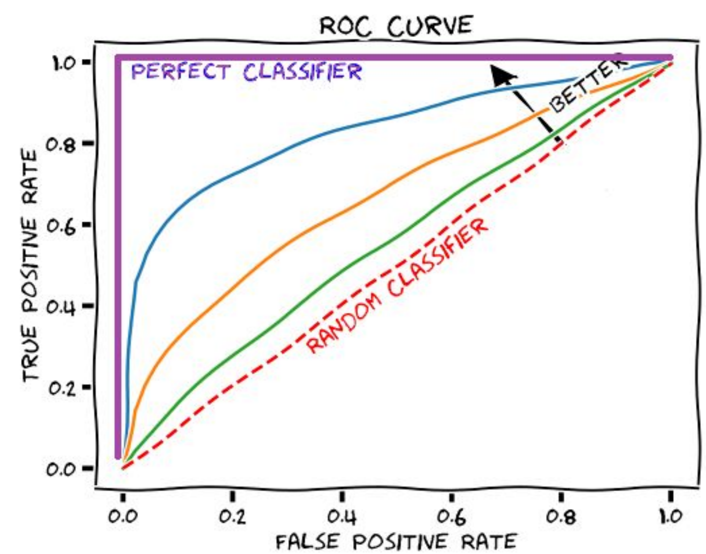
- Coefficient of determination (R^2)
- Max. Absolute Error (MaxAE)
- Mean Absolute Error ± Standard Deviation (MAE ± STD)
- Root Mean Square Error (RMSE)

3. Evaluation Metrics and Protocols

Classification metrics:

- Precision and Recall
- F1-Score
- Balanced Accuracy
- Confusion Matrix Plot
- ROC AUC Curve

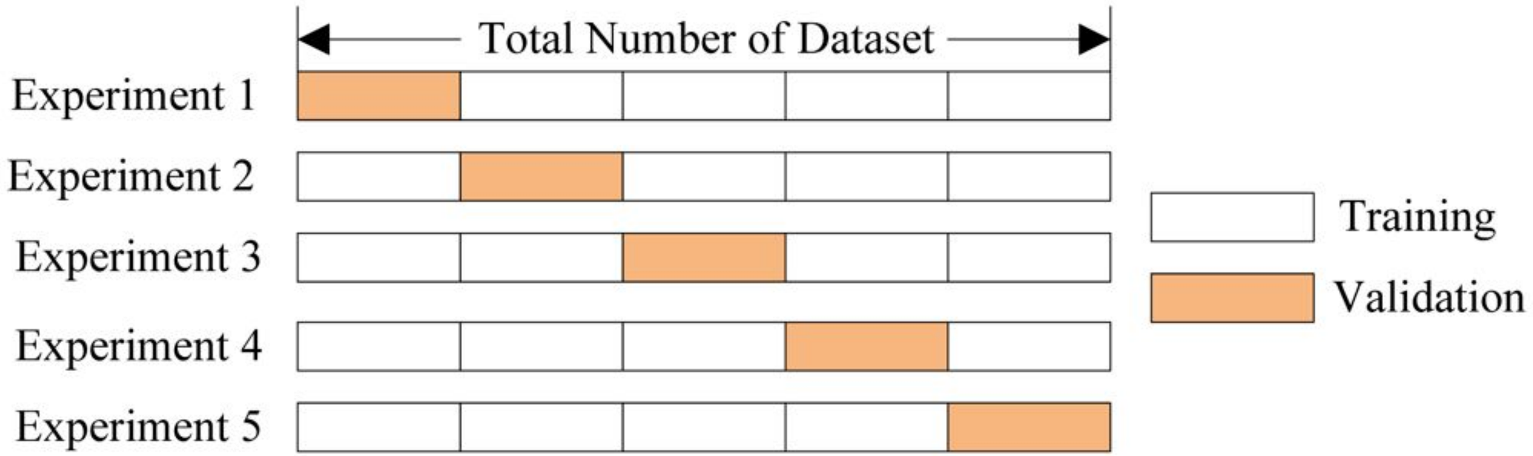




3. Evaluation Metrics and Protocols

K-Fold cross-validation:

- For a better generalization and confidence in the model
- K = 5 → 20% each set



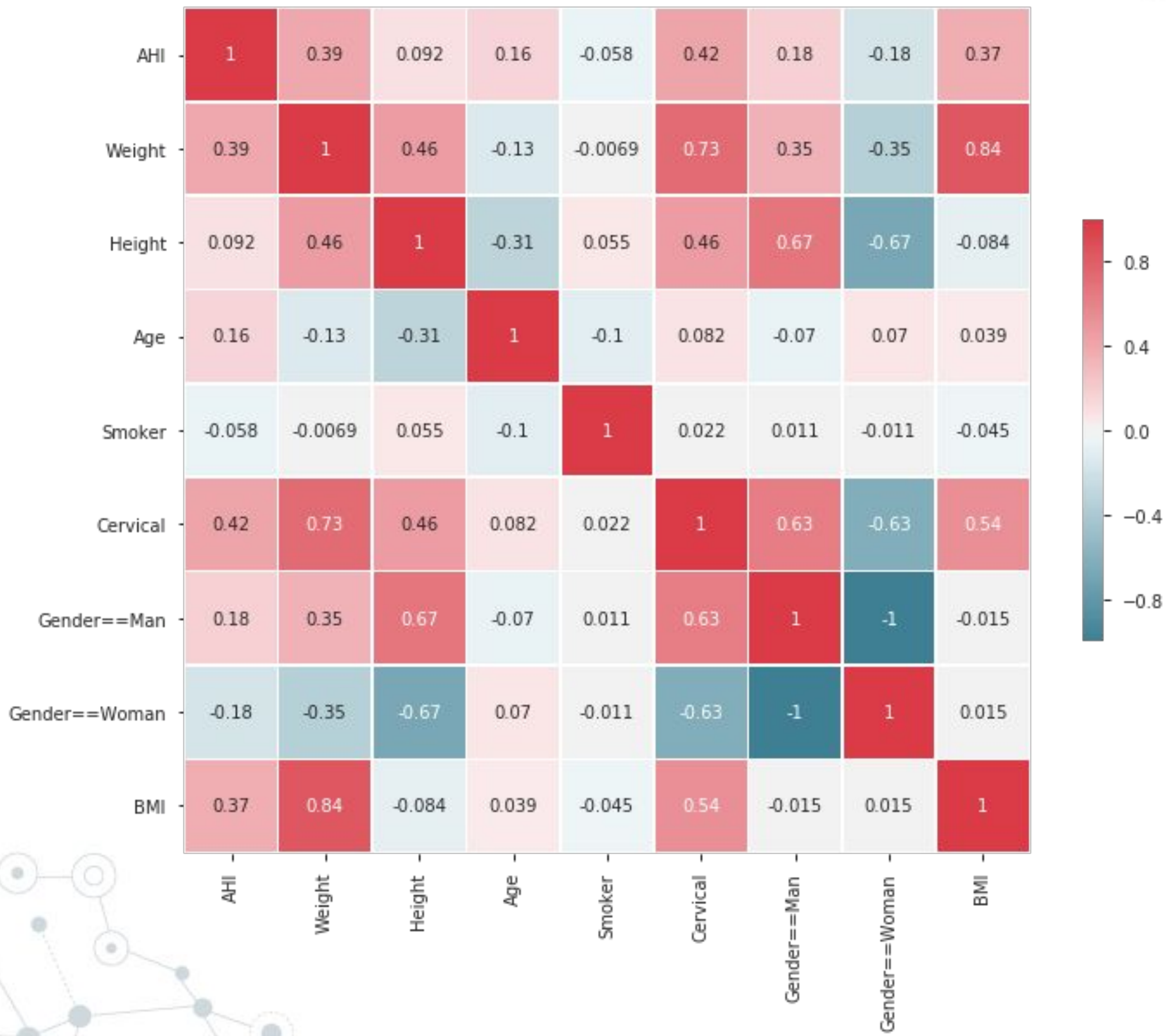
4.

Exploratory Data Analysis

Approach to analyze datasets: summarize their main characteristics, discover patterns, spot anomalies, test hypothesis… Know your data!
Often with visual methods

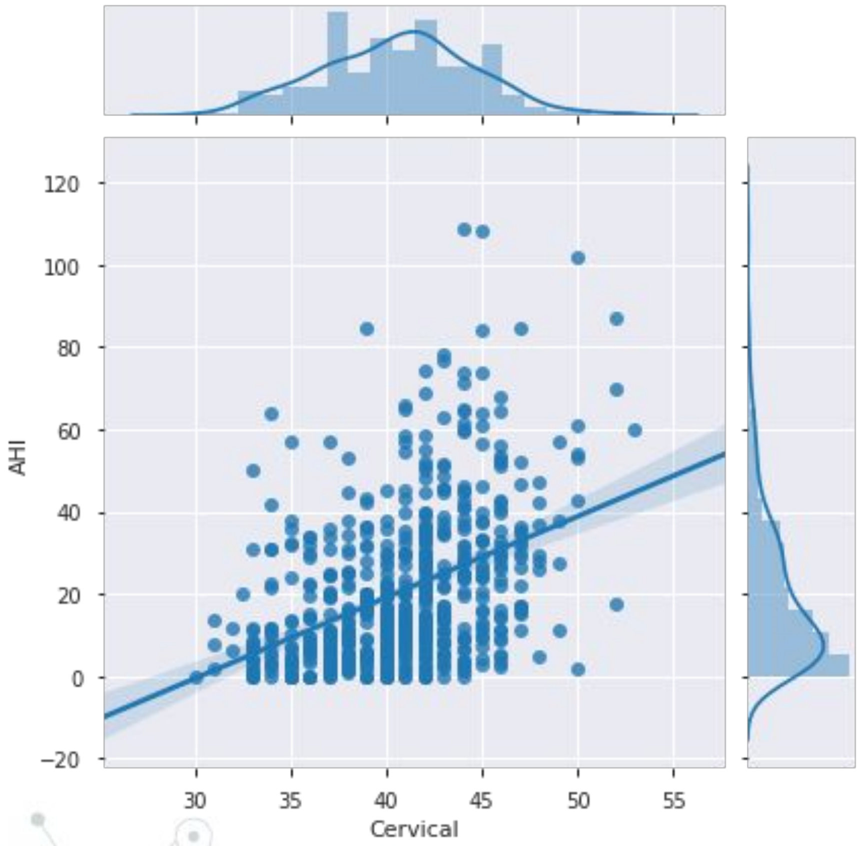
4. Exploratory Data Analysis

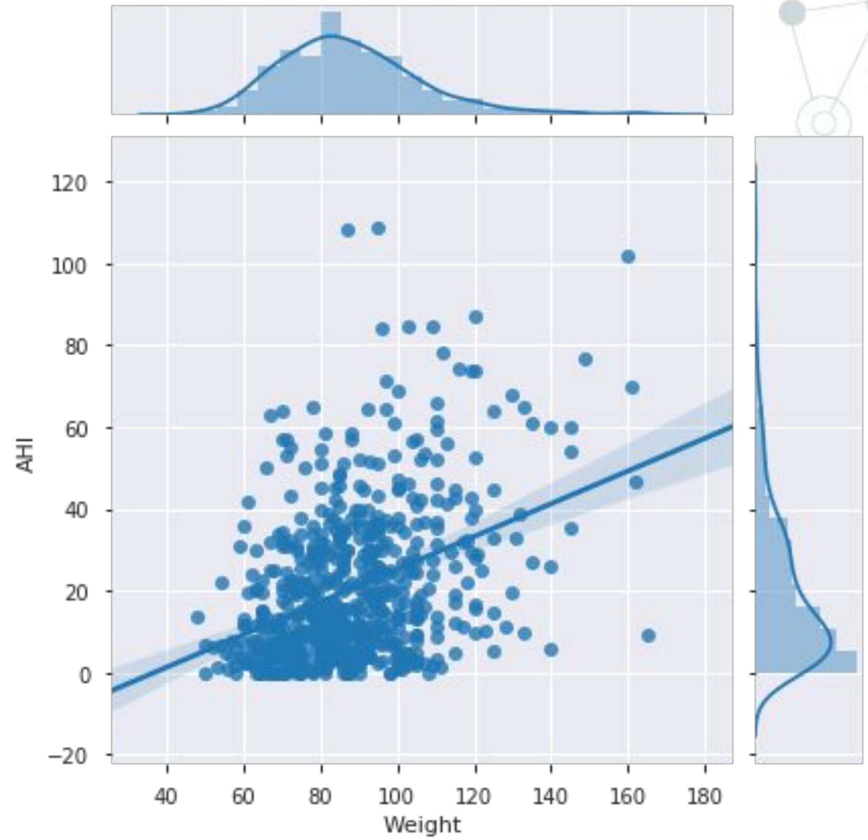
Correlation Matrix:



4. Exploratory Data Analysis

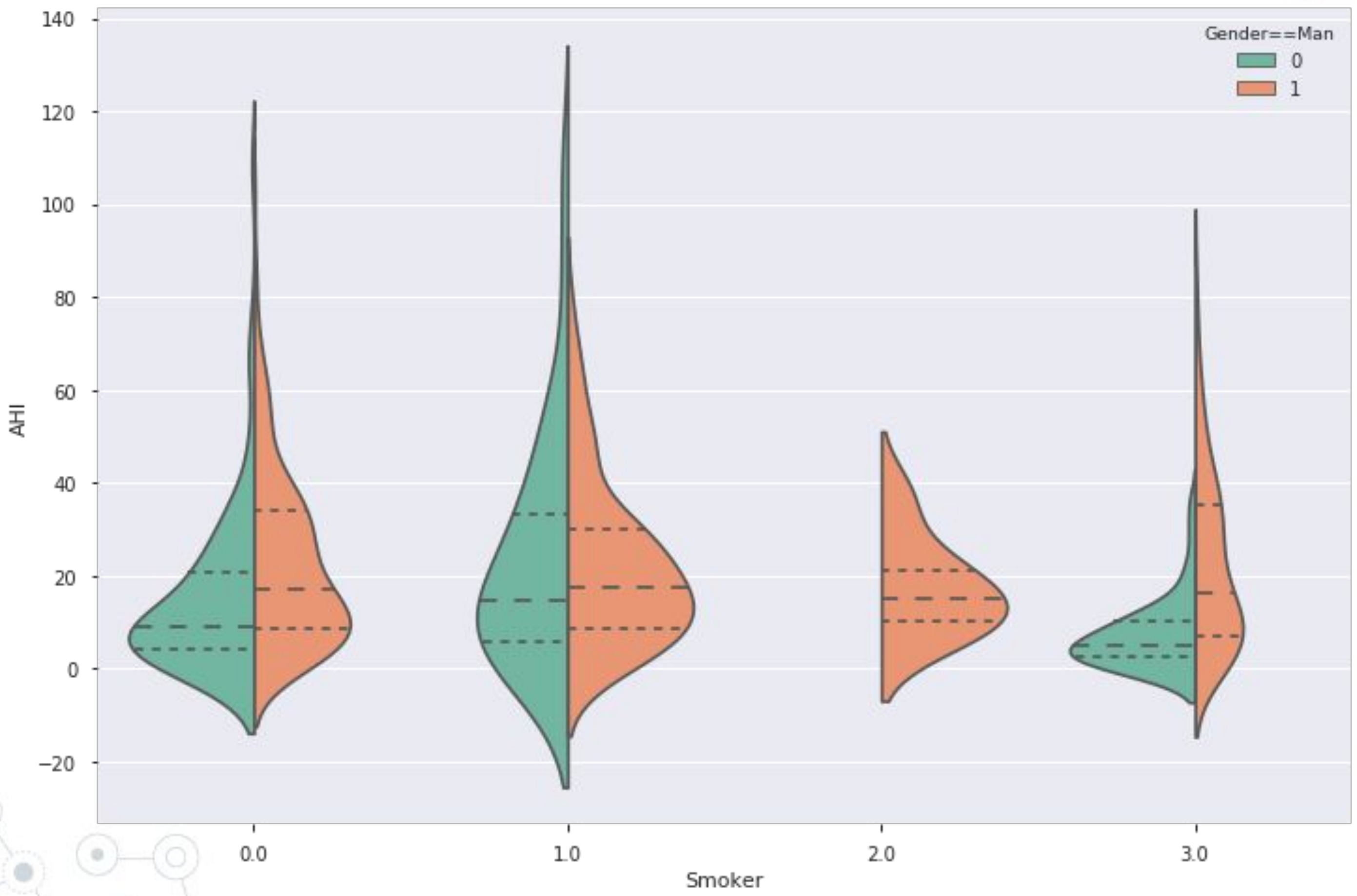
Join Plots:





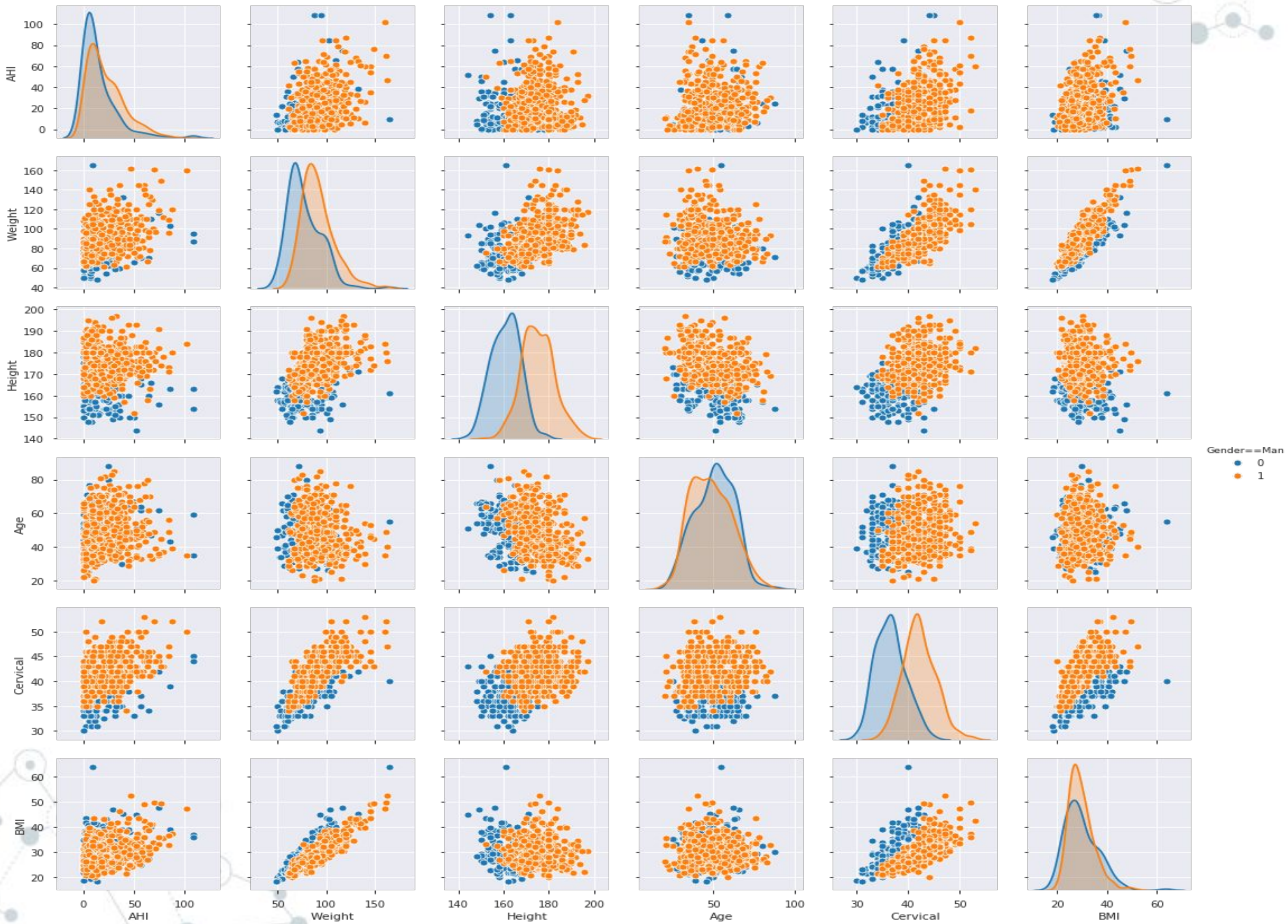
4. Exploratory Data Analysis

Violin Plot (categorical vars ⇔ target):



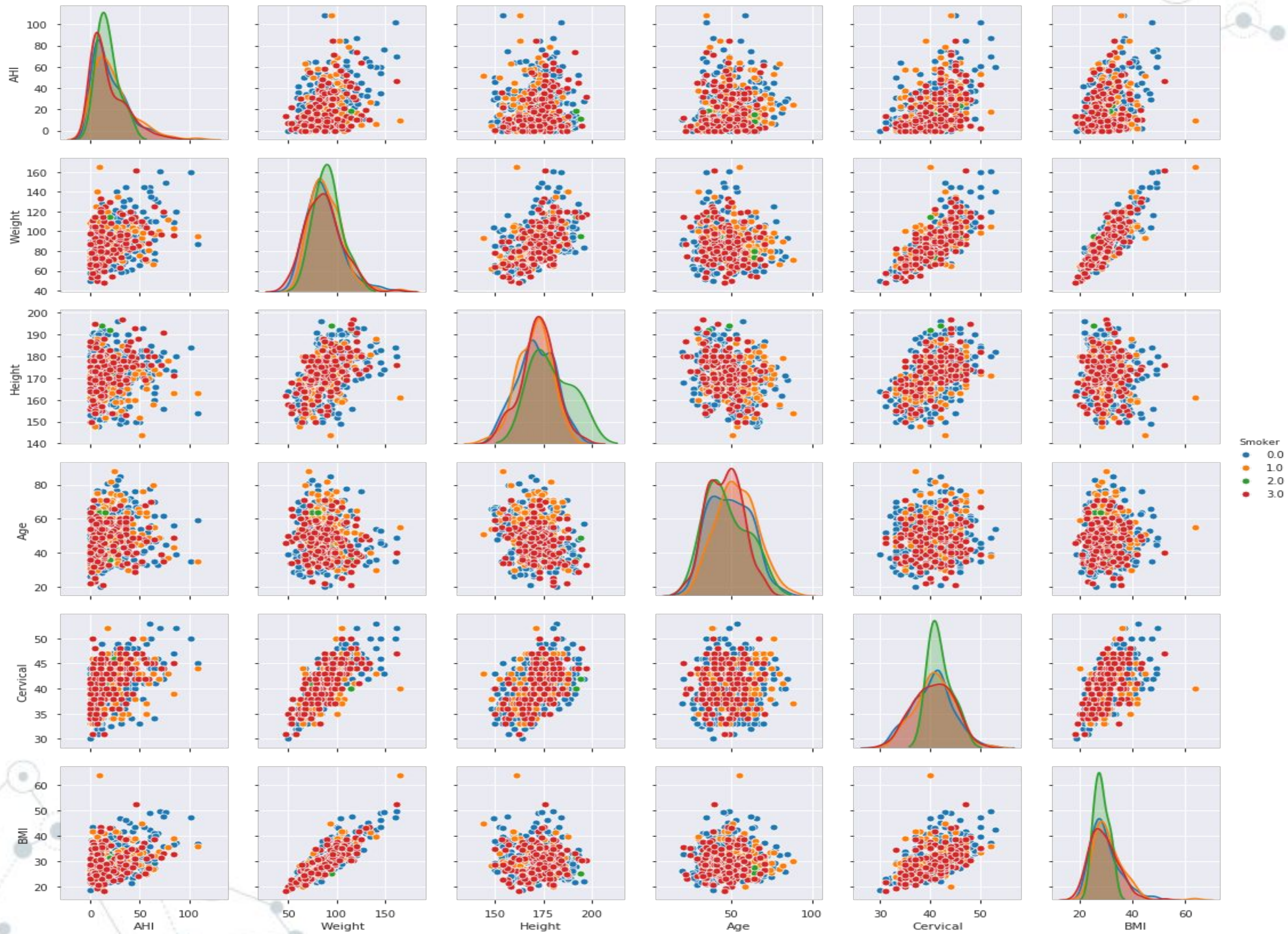
4. Exploratory Data Analysis

Pair Plots (hue Gender):



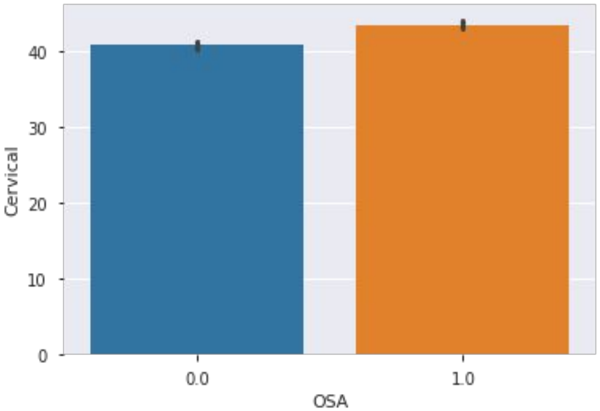
4. Exploratory Data Analysis

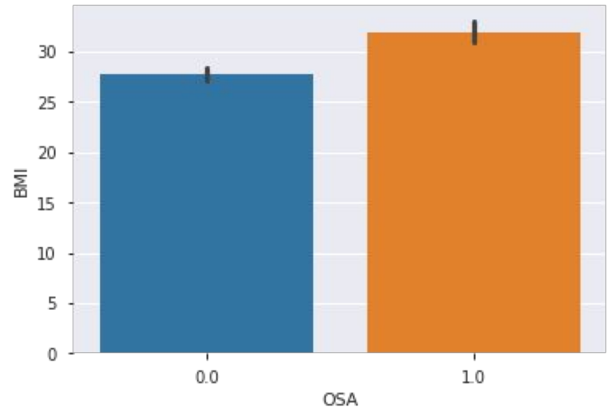
Pair Plots (hue Smoker):

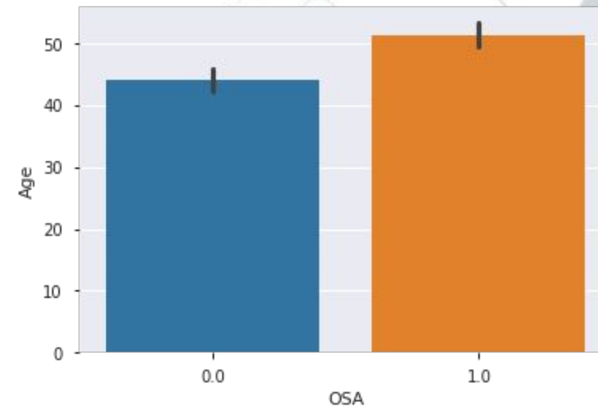


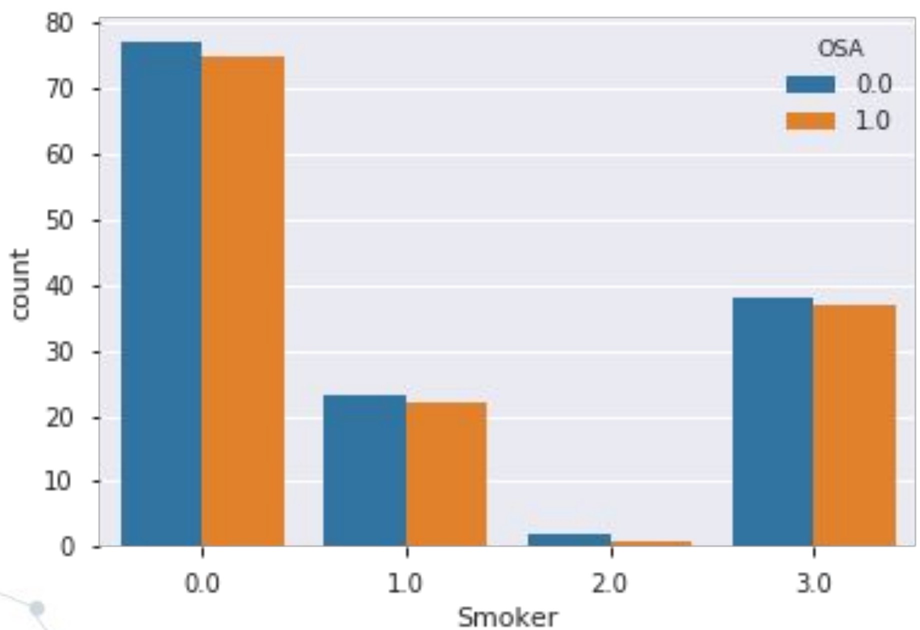
4. Exploratory Data Analysis

Classification:









5.

Model Selection and Training

Implementing the Machine Learning models

5. Model Selection and Training

Data Preprocessing:

- Polynomial Features

$$[a, b] \rightarrow [1, a, b, a^2, ab, b^2]$$

- Standard Scaler

$$z = \frac{x - \mu}{\sigma}$$

μ = Mean
σ = Standard Deviation

- MinMax Scaler

$$z = \frac{x - \min(x)}{\max(x) - \min(x)}$$

5. Model Selection and Training

Implemented Models: Why not all of them?

- Generalized Linear Models
- Support Vector Machines
- Nearest Neighbors
- Gaussian Processes
- Decision Trees
- Ensemble Methods
- XGBoost and CatBoost (Gradient Boosting Decision Trees)
- Neural Networks

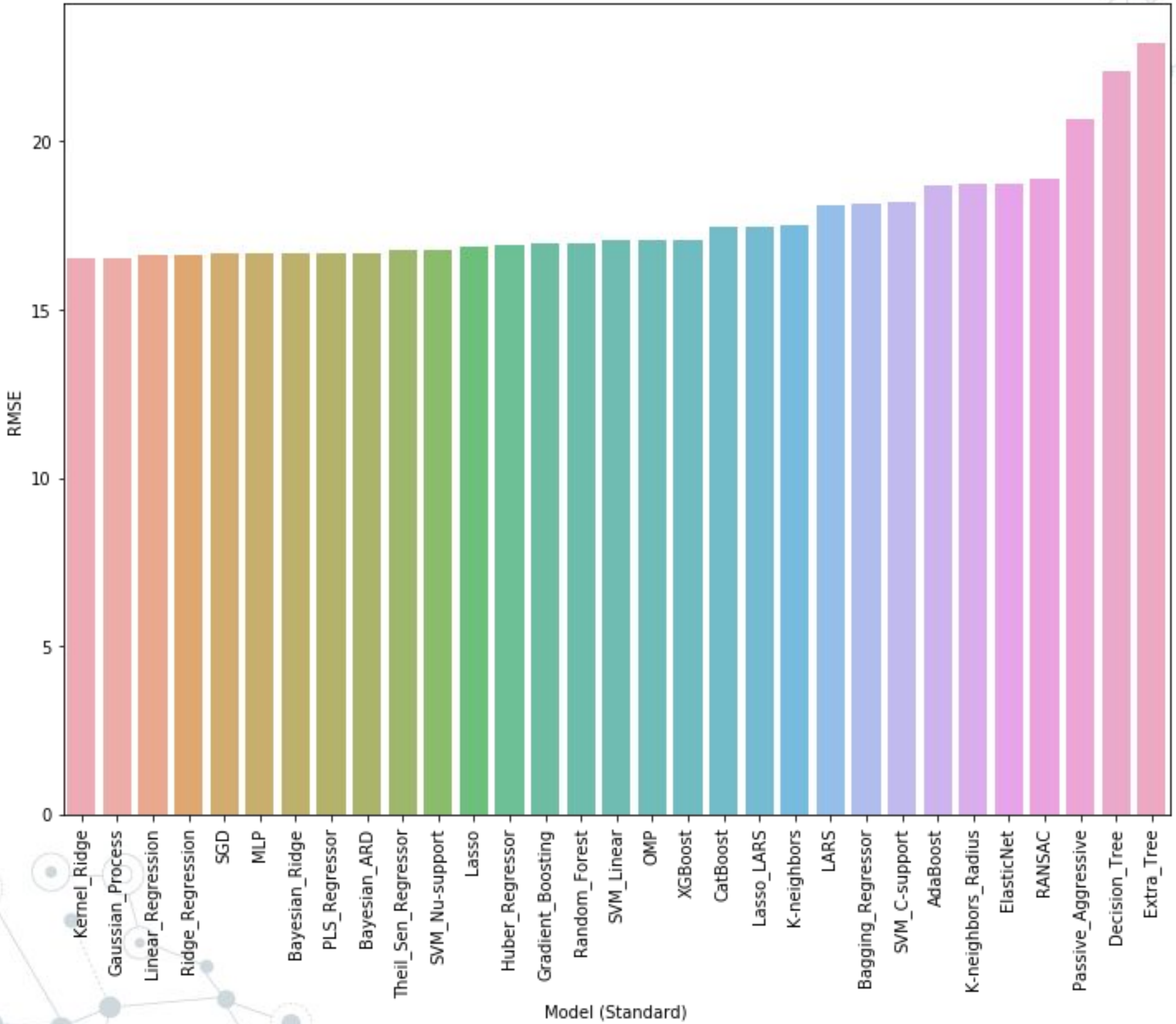
6.

Model Testing and Results

Comparing the performance of the models

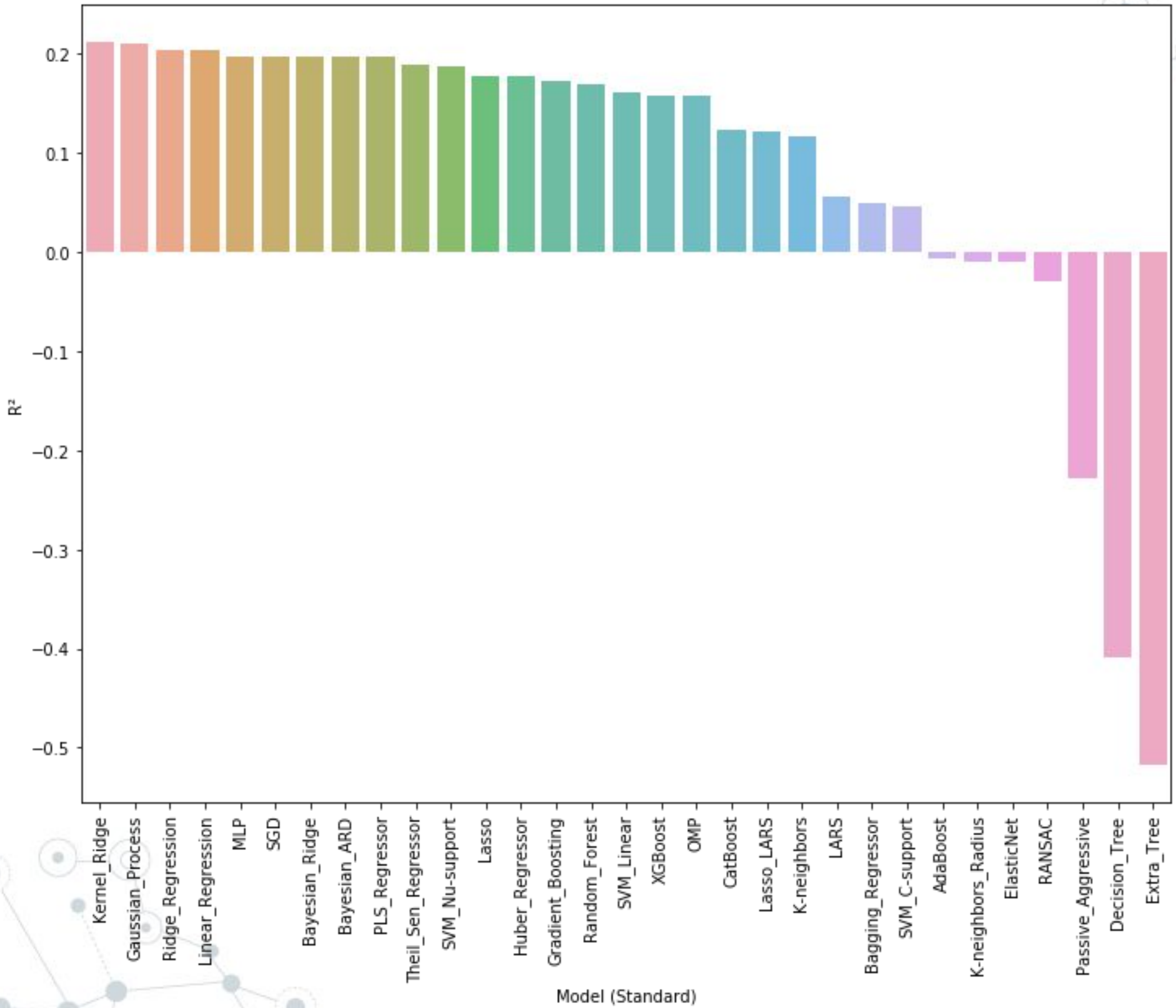
6. Model Testing and Results

Regression Models: (Standard Scaling)



6. Model Testing and Results

Regression Models: (Standard Scaling)

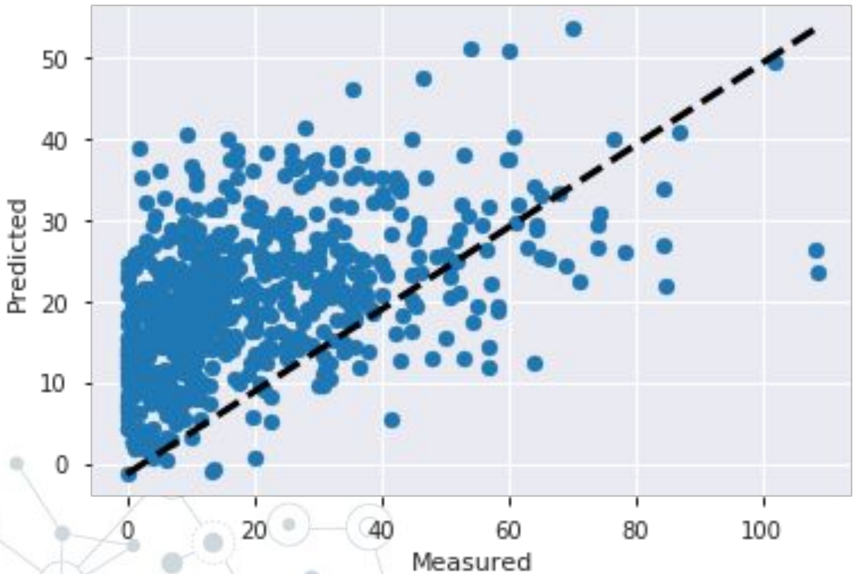


6. Model Testing and Results

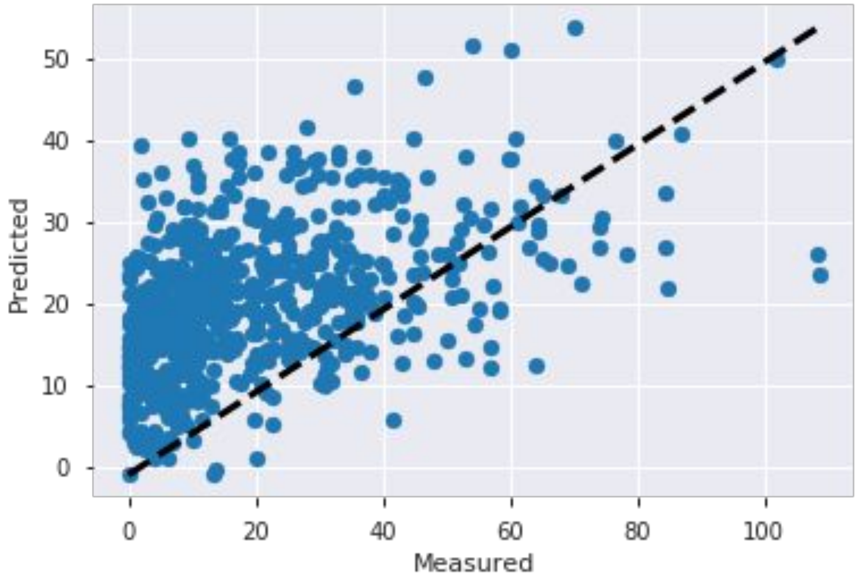
Regression Models: (Standard Scaling)

Model	R^2	Max Error	MAE ± STD	RMSE
Kernel_Ridge	0.213	84.91	12.49 ± 16.53	16.53
Gaussian_Process	0.211	84.99	12.50 ± 16.54	16.54

Kernel Ridge



Gaussian Process



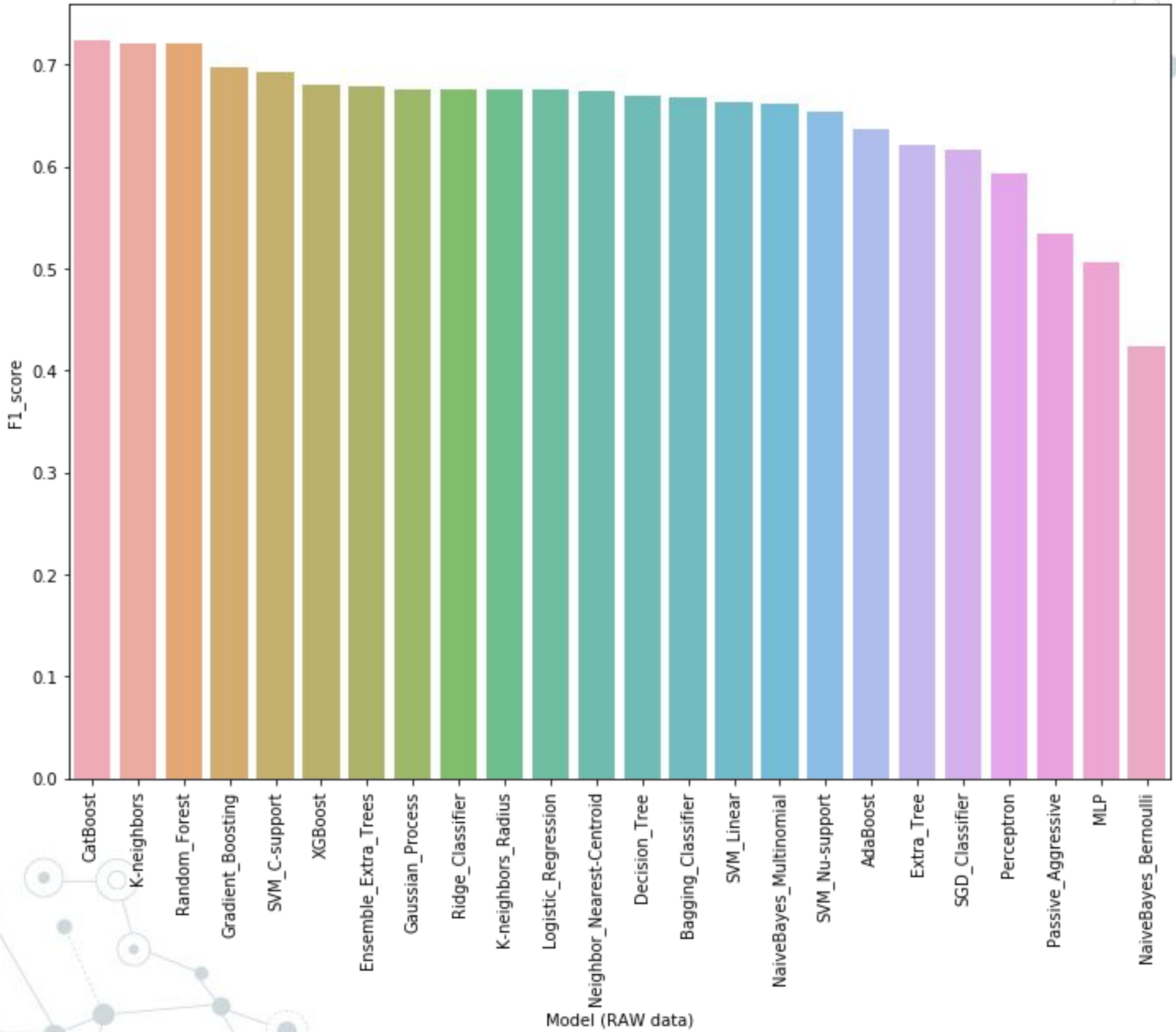
6. Model Testing and Results

Regression Models Discussion:

- No sufficient precision to be useful in real-world
- Despite a certain correlation between real and predicted values, the variance is very high and there are very remarkable mistakes
- The positive aspect… Best models are *white box* (explainability)
- The most complex models are not leading the results… This may indicate that the weak results are not caused by a bad choice of hyperparameters or models
- The problem may be in the data itself (lack of samples or too high complexity of the problem to be solved with the used variables)

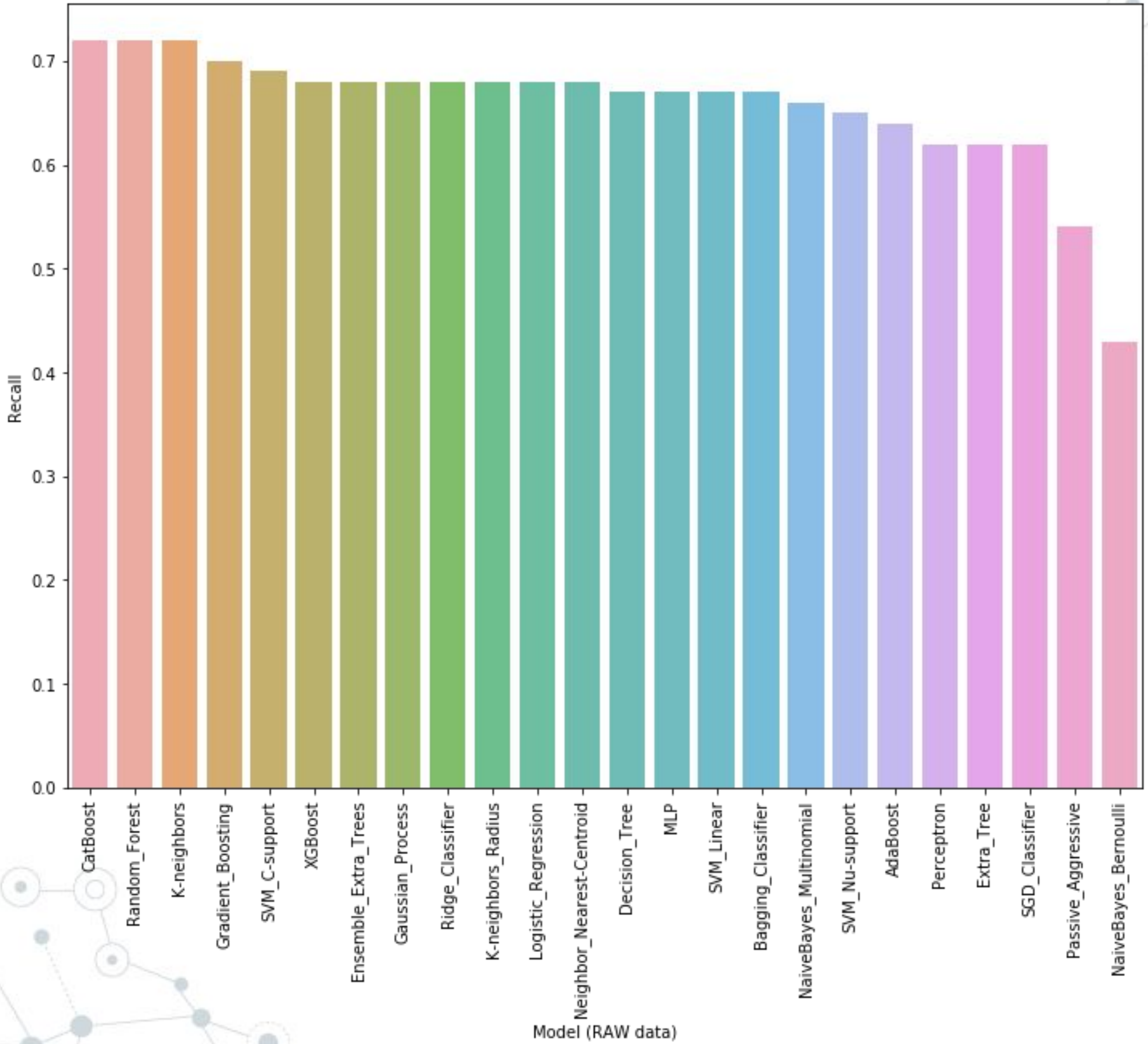
6. Model Testing and Results

Classification Models: (Raw Data)



6. Model Testing and Results

Classification Models: (Raw Data)



6. Model Testing and Results

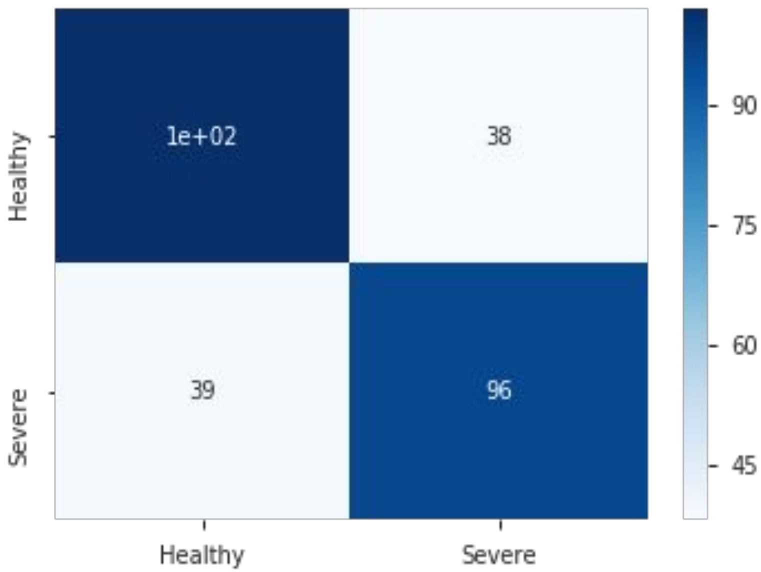
Classification Models: (Raw Data)

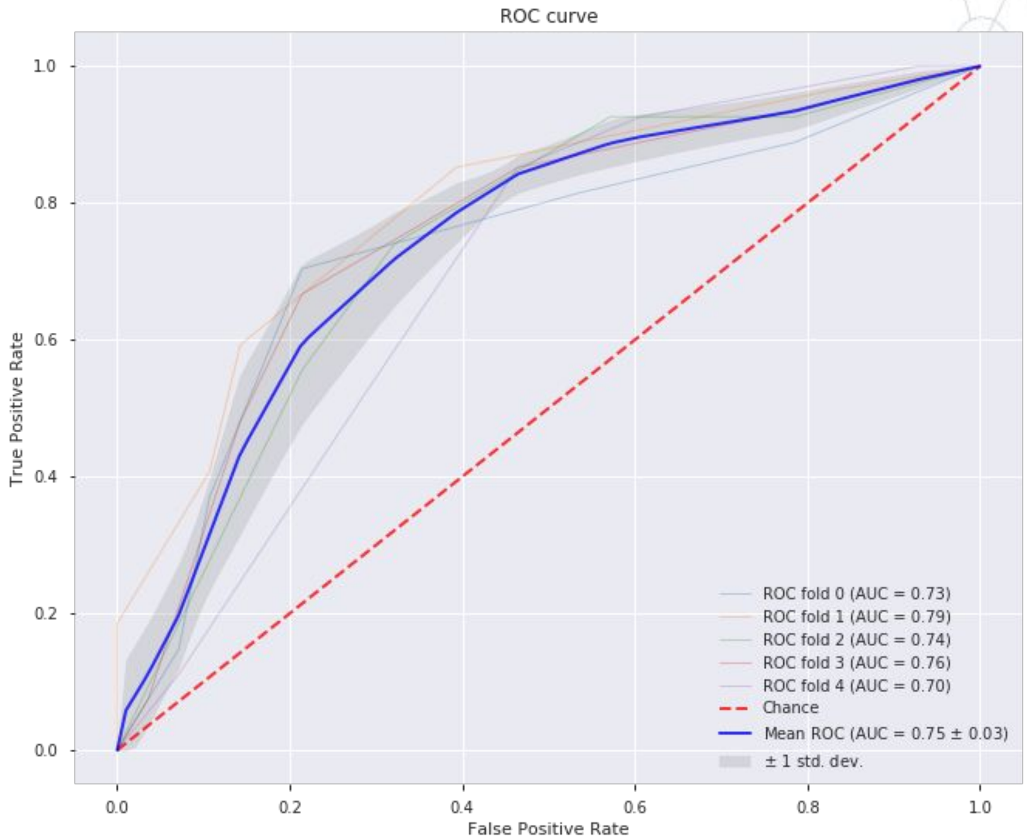
Model	Precision	Recall	F1-Score	Balanced Accuracy
K-neighbors	0.72	0.72	0.720	0.720
Random_Forest	0.72	0.72	0.720	0.720
CatBoost	0.72	0.72	0.724	0.724

6. Model Testing and Results

Classification Models: (Raw Data)

K-Neighbors

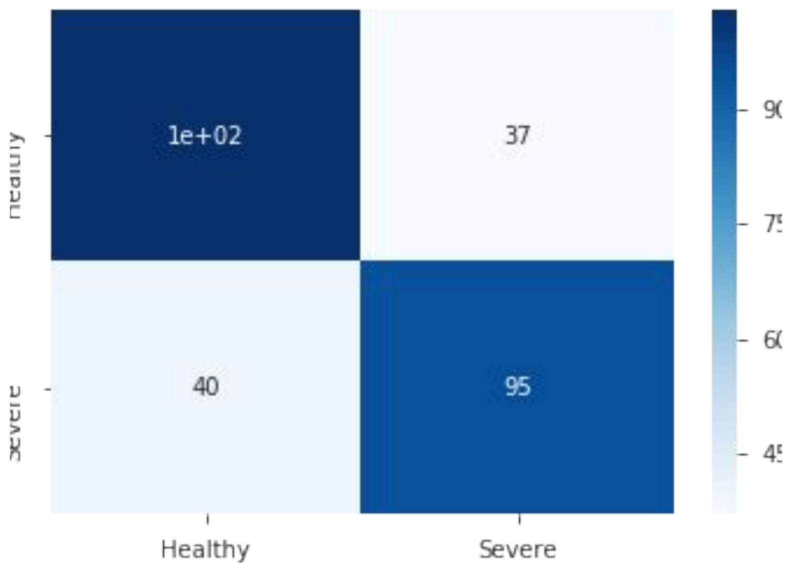


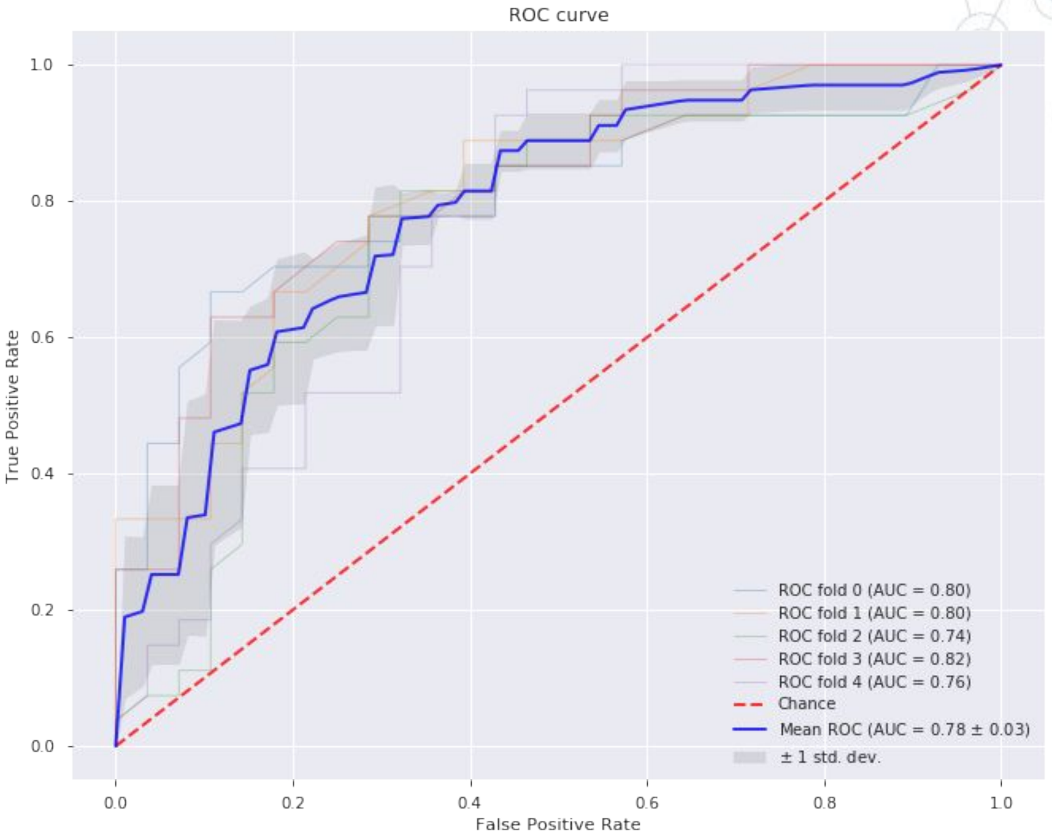


6. Model Testing and Results

Classification Models: (Raw Data)

Random Forest

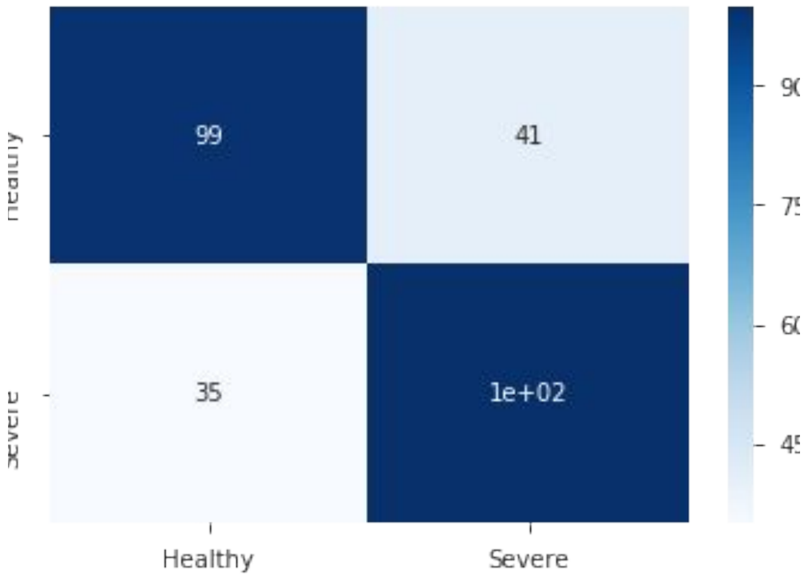


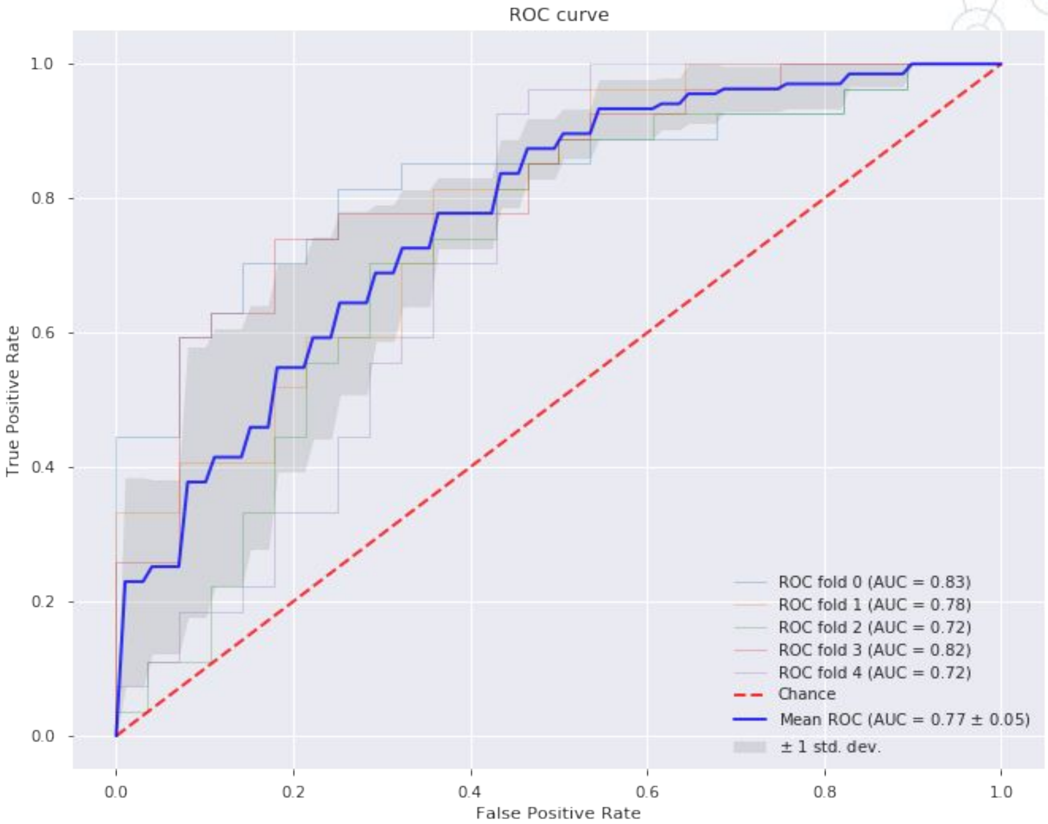


6. Model Testing and Results

Classification Models: (Raw Data)

CatBoost





6. Model Testing and Results

Regression Models Discussion:

- Certainly good results (easier problem than regression…)
- Best results obtained with *CatBoost* model, but...
- In health applications minimizing False Negatives (FN) is critical → Minimize severe patients classified as healthy → *Random Forest* model
- Explainability is also very important in this field → *K-Neighbors* model

Conclusions

What lessons can we draw from this project?

Conclusions

- In regression approach, it has not been possible to obtain models good enough for deployment in real-world scenarios
- In classification approach, certainly good results
- This weak results are probably due to data itself (lack of samples or too high complexity of the problem to be solved with the used variables)
- Python and the presented libraries for Machine Learning problems and data processing, provides a very powerful and user-friendly framework for ML development

Future Lines

What can be done to improve the models?

Future Lines

- Use of the frequencies audio dataset (Feature Selection)
- Feature Engineering (linear combinations of the variables)
- Methods for handling with missing values (mean, median, last/next observation, interpolation, most frequent value…)
- Different scalers for the input data
- Ensembling methods to combine multiple models
- Hyperparameter optimization on the best models (grid search, random search, bayesian optimization…)
- Analysis of the requirements (computational and time) of the models used, since is a critical factor for a real-world deployment

Thanks!

Any questions?

You can find me at:

GitHub www.github.com/jaimeperezsanchez/

www.linkedin.com/in/jaime-perez-sanchez/

jaime.perez.sanchez@alumnos.upm.es