# VITAL SIGNALS FOR EARLY MORTALITY PREDICTION IN ICU PATIENTS

RENATO DIAS, NMEC: 98380

DATA PREPARATION

- Remove commas and convert each value from object type to float type;
- Drop missing values by eliminate the sample that contains the 2 missing values;
- Other less significant changes like show full numbers instead of number in scientific notation.

Target Variable Distribution
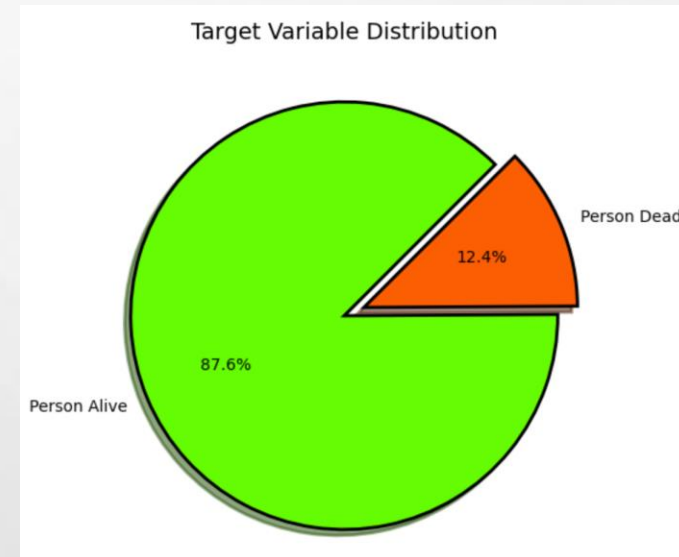
	count
max	2725.00
min	2725.00
mean	2725.00
median	2725.00
mode	2725.00
std	2725.00
var	2725.00
range(max-min)	2725.00
kurtosis	2724.00
skewness	2724.00
energy	2725.00
periodogramPower	2725.00
Alive	2725.00

MODELS OVERVIEW

- 8 MACHINE LEARNING MODELS: 4 TRANSPARENT CLASSIFIERS (DECISION TREE, LINEAR DISCRIMINANT, LOGISTIC REGRESSION, LINEAR SVM) AND 4 NON-TRANSPARENT MODELS (RANDOM FOREST, K-NEAREST NEIGHBORS, BOOSTED TREE AND GAUSSIAN SVM)

- *EVAL_METRICS* FUNCTION TO BUILD AND EVALUATE THE MODELS:

  - ACCURACY

  - PRECISION

  - RECALL

  - F1 SCORE

  - AUC

TUNNING HYPERPARAMETERS

- *GridSearchCV* OBJECT TO TUNE THE HYPERPARAMETERS OF EACH MODEL, WHICH TAKES 3 ARGUMENTS:
  - THE MODEL THAT WE WANT TO TUNE;
  - THE HYPERPARAMETERS THAT WE WANT TO TUNE;
  - NUMBER OF FOLDS TO USE FOR CROSS-VALIDATION.

- HYPERPARAMETERS TUNNED:
  - C
  - MAX_DEPTH
  - N_ESTIMATORS
  - CRITERION
  - N_NEIGHBORS
  - LEARNING_RATE
  - SOLVER

RESULTS

Models Evaluation:

Model Name	Accuracy Score	Precision	Recall	F1-score	AUC
Decision Tree Classifier	0.76	0.88	0.85	0.86	0.48
Random Forest Classifier	0.88	0.88	1.00	0.94	0.50
K-Nearest Neighbors Classifier	0.86	0.88	0.97	0.92	0.49
Linear SVM	0.88	0.88	1.00	0.94	0.50
Gaussian SVM	0.88	0.88	1.00	0.94	0.50
Logistic Regression	0.88	0.88	1.00	0.94	0.50
Linear Discriminant	0.88	0.88	1.00	0.94	0.50
Boosted Tree	0.87	0.88	0.99	0.93	0.49

Models Evaluation:

Model Name	Accuracy Score	Precision	Recall	F1-score	AUC
Decision Tree Classifier	0.88	0.88	1.00	0.94	0.50
Random Forest Classifier	0.88	0.88	1.00	0.94	0.50
K-Nearest Neighbors Classifier	0.88	0.88	0.99	0.93	0.50
Linear SVM	0.88	0.88	1.00	0.94	0.50
Gaussian SVM	0.88	0.88	1.00	0.94	0.50
Logistic Regression	0.88	0.88	1.00	0.94	0.50
Linear Discriminant	0.88	0.88	1.00	0.94	0.50
Boosted Tree	0.88	0.88	1.00	0.94	0.50

CONCLUSION

- THIS PROJECT SUGGEST THAT MACHINE LEARNING IS A POWERFUL TOOL THAT CAN BE USED TO SOLVE A VARIETY OF PROBLEMS, BUT IT IS IMPORTANT TO EVALUATE THE PERFORMANCE OF MACHINE LEARNING MODELS TO ENSURE THAT THEY ARE WORKING AS EXPECTED;

- THERE ARE A VARIETY OF EVALUATION METRICS THAT CAN BE USED TO ASSESS THE PERFORMANCE OF MACHINE LEARNING MODELS;

- DATA VISUALIZATION CAN BE HELPFUL FOR UNDERSTANDING THE PERFORMANCE OF MACHINE LEARNING MODELS.