LUMEN ET VIRTUS

Classification model for prediction of death from heart failure

https://doi.org/10.56238/levv15n39-075

Daniel Baldini Filipe[1] and José Erasmo Silva[2]

ABSTRACT

Heart failure is a syndrome in which the heart is unable to pump blood at optimal levels throughout the body. In recent decades, there has been a significant increase in deaths caused by this condition. This growth is associated with changes in behavior and aging, among other factors. Given the growing challenge presented by this scenario, it is essential to develop a sophisticated classification model that can predict death and that is easily integrated into existing health systems. In light of this challenge, five classification models were created for a sample of 299 patients and their results were compared. This sample contains 12 explanatory variables, in addition to the response variable, which indicates whether the patient died during treatment. The results achieved were satisfactory and show that, although more than one model presents good results, considering technical and analytical aspects, the binary logistic model is the one that presented the best balance of metrics. Because it is a high-performance, easy-to-interpret algorithm with low cost of training, availability and execution, it is concluded that the binary logistic model created in this study can be integrated into current systems in order to help health professionals with their diagnosis and, thus, they can suggest changes in habits and/or treatment to the patient so that he or she can have a longer and healthier life.

Keywords: Health, Heart Problem, Machine Learning, Data Analysis, Data Balancing.

[1] Data Science and Analytics Specialist
University of São Paulo
[2] Doctor in Business Administration
Federal University of Bahia

INTRODUCTION

In the 1950s, research began to create what we now know as "machine learning". In 1957, Frank Rosenblatt created the first "software" capable of learning as it received data over time (Taulli, 2020). His idea was based on the human neuron and became known as Rosenblatt's "perceptron". However, mainly due to the computational limitations of the time, it was only possible to create neural networks with a "perceptron", which made this topic forgotten for a few decades.

In the 1980s, Rosenblat's ideas came back to the fore and this brought about what some authors call the Artificial Intelligence [AI] revolution. With computational power far superior to those of the 1950s, Rosenblatt's concept of the "perceptron" was expanded to give rise to what we now call "deep learning."

The use of machine learning in healthcare has been growing significantly in recent years. During the COVID-19 pandemic, these algorithms proved to be effective and important both for the development of the vaccine (Park, 2021) and for the management of hospital materials and equipment that were so scarce in that period.

The adoption of machine learning techniques for detecting cardiovascular diseases has been gaining ground, as it is estimated that 620 million people live under some disease in this group and that it is responsible for approximately 17.9 million deaths per year. According to a study published by the British Heart Foundation, this number is expected to increase in the coming years due to factors such as: changes in behavior, growth and aging of the world population, and growth in the survival rate after heart attacks and strokes (BHF, 2023; WHO, 2023).

In order for these techniques to be able to adequately predict and control the onset or evolution of cardiovascular diseases, it is crucial to identify and monitor risk factors such as hypertension, cholesterol, obesity, among others (Uddin et al., 2022). Visible signs such as gross weight gain, development of edema, ankle swelling should also be considered relevant factors and be monitored so that the necessary measures are taken in time, thus preventing the evolution to more serious situations (Escolar et al., 2021).

Part of the group of cardiovascular diseases, heart failure is a syndrome in which the heart is unable to pump blood in order to supply all the metabolic needs of the tissues with the appropriate amount of blood (Rohde et al., 2018). According to Ferreira (2020), it can also generate swelling or congestion due to the accumulation of blood in a region caused by the reduction of its flow.

According to Arruda et al. (2022), the period between 1998 and 2019 in Brazil was marked by a significant number of deaths related to heart failure, especially in adults over 50 years of age. During those years, the total number of deaths attributed to this condition reached 567,789. This data highlights the growing concern about heart failure as an important cause of mortality in the country's older population. The trend observed over these two decades suggests the need for greater attention

to heart health and the management of chronic diseases, especially in the elderly, to combat this growing cause of mortality.

In view of the growing concern about heart disease in Brazil, especially motivated by the aging of the population, there is a need to explore research methods that offer information on the risk factors associated with cardiac events. In this context, binary and multinomial logistic regression models are used by researchers who are interested in evaluating the probability of infarction in a group of people based on physical characteristics such as: weight, waist circumference, and also on their health habits such as: eating habits, physical exercise, smoking, among others. For this study, infarction is the variable to be predicted and the event may or may not occur due to the explanatory variables provided to it. Thus, infarction is a dichotomous qualitative variable and, as the intention is to estimate the probability of occurrence of this event, binary logistic regression can be used (Fávero and Belfiore, 2017).

Therefore, in the face of the growing challenge posed by heart failure, it is essential to develop a sophisticated classification model using advanced machine learning techniques. This model should be able to analyze and interpret a variety of clinical and behavioral data – similar to those used in the binary logistic regression example of Fávero et al. (2017) – to accurately predict the probability of death caused by heart failure.

Additionally, it is crucial that this model is designed to efficiently integrate with existing healthcare systems, allowing healthcare professionals to utilize their predictions to make more informed clinical decisions. By doing so, the model will not only contribute to the earlier detection and more effective treatment of heart failure, but will also help to prolong and improve the quality of life for patients. The upward trend in deaths from heart failure, as highlighted by Arruda et al. (2022), emphasizes the urgency of such an innovation. Thus, the goal of this work is to develop a classification model that not only advances the field of machine learning in healthcare, but also has a significant and practical impact on the management and treatment of cardiovascular diseases.

MATERIAL AND METHODS

The database used for this study was obtained from the following address: *https://archive.ics.uci.edu/dataset/519/heart+failure+clinical+records.* It contains the medical records of 299 patients and has the following variables: age, sex, whether the patient is a smoker, whether he has diabetes, anemia and high blood pressure high_blood_pressure). In addition to these best-known data, it also brings: the amount of platelets ("platelets") which are the blood cells responsible for blood clotting; The ejection fraction ("ejection_fraction") which is the percentage of blood ejected by the heart with each beat; Serum creatinine ("serum_creatinine") which is a chemical residue produced by the death of muscle cells; Creatinephrophosphokinase

("creatinine_phosphokinase") which is an enzyme present in muscle tissues, the heart and the brain; Serum sodium ("serum_sodium") which brings the amount of sodium in the patient's blood and, finally, the time variable ("time") which brings the number of days the patient was followed by the doctor. The last piece of information is the response variable, "death_event", which determines whether the patient died during treatment.

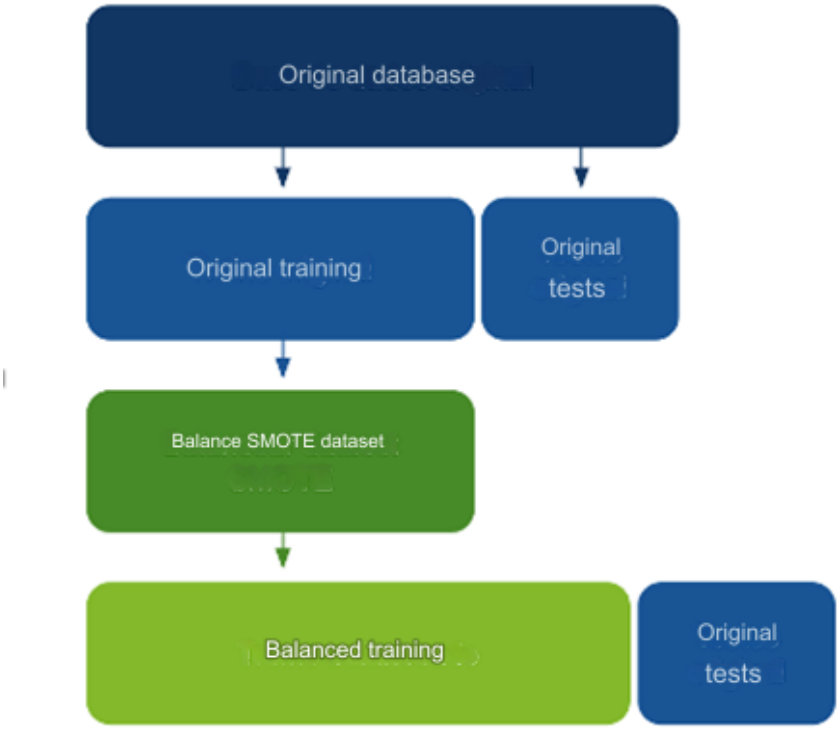
It did not contain any non-existent value in the data and no processing or alteration of the information was necessary. The variables "anaemia", "diabetes", "high_blood_pressure", "sex", "smoking" and "death_event" were converted from the integer type to the factor type and the variable "age" was converted to integer.

In addition to adjusting the models with the original database, this study also used the Synthetic Minority Oversampling Technique (SMOTE) to add observations to the minority category in the training base and adjust the models from the new dataset.

SMOTE is considered the most prominent method to circumvent the frequent issue of unbalanced data, because instead of replicating data already existing in the database, it uses linear interpolation of minority class data and techniques such as "K nearest neighbors" to create synthetic examples of the minority class (Elreedy et al., 2022) which can help in generalizing the model and also avoid its "overfitting".

The test base was not balanced in any scenario, so only observations existing in the original database remained, as shown in Figure 1.

Figure 1. Data preparation flow for balanced-based models



Source: Original survey data

In search of the most appropriate model for predicting the response variable, the following models were implemented, predicted and analyzed: Binary logistic regression, vector machine and support, decision tree, gradient boosting and naïve Bayes.

To facilitate the reproduction of the results, the value 2024 was used as a seed in all stages of this study. Determining a seed is a technique used to ensure that the results obtained are always the same, as this seed ensures that all executions have the same randomness.

LOGISTIC REGRESSION

The logistic model was developed at the end of the nineteenth century, but became popular from the second decade of the twentieth century.

In addition to being widely used in the health area, it is also effective for forecasts in the financial, insurance, and other areas to define, for example, whether or not credit should be granted, or even what is the level of risk of a certain customer suffering an accident (Fávero and Belfiore, 2017).

In this study, the "stepwise" procedure was used, as it maintains only the predictive variables necessary for the chosen level of significance.

The main advantages of this model are: it requires low computational performance and is easy to implement and interpret. Its main limitations are related to multicollinearity and/or heteroscedasticity in the explanatory variables.

SUPPORT VECTOR MACHINES

In the early 1990s, as an alternative to artificial neural networks for nonlinear classification and regression tasks, the technique of support vector machines was developed and gained relevance because it has a greater capacity for generalization even though it is estimated from a database without many records (Gholami and Fakhari, 2017).

To estimate this model, the standardization of the variables using the "Z-Score" technique was used: eq. (1). This technique makes the standardized variables have a mean of 0 and variance of 1 using the form. (1) (Fávero and Belfiore, 2017).

$$Z = \frac{X - \mu}{\sigma} \qquad (1)$$

where, Z: is the index; X: is the amount that needs to be converted; μ: is the average of all X values; σ: is the standard deviation of the sample.

In addition to Z-Score, this model also made use of cross-validation. According to Guerbai et al. (2022), this technique consists of separating the data into "n" subsets of the same size. The

algorithm then selects n-1 parts for estimation and the part not considered is used for testing. This operation is repeated until all parts have been used for both estimation and testing. The accuracy of the model is calculated at the end of this iterative process.

The main advantages of support vector machines are: low noise influence on the data, being able to classify nonlinear problems and being effective on large volumes of data. The main disadvantages are the difficulty of interpreting and visualizing the hyperplane, being an algorithm that requires greater computational power to be executed, compared to logistic regression and the possibility of "overfitting″ due to the choice of hyperparameters (Boswell, 2002).

DECISION TREES

It is a powerful and versatile algorithm capable of predicting categorical and continuous data. From sequential tests, it compares an attribute to a threshold, adding the observation to the most frequent class. As it is the fundamental component of "random forests", it is possible to say that it is one of the most important algorithms today (Géron, 2022; Kotsiantis, 2011).

When the data does not have many variables and/or does not require many branches in the tree to reach the prediction, it is a model that is easy to implement, interpret and quickly execute, otherwise this model becomes difficult to interpret, the speed of execution is affected and it can also incur in "overfitting" (Sosnovshchenko et al., 2018).

GRADIENT BOOSTING

It is an advanced technique for predictive modeling in structured data, employing the strategy of "gradient boosting with component-to-component linear models" (glmboost), an approach that is distinguished by its focus on sequential and incremental model construction. Unlike traditional gradient boosting, which often uses decision trees as base models to reduce residuals and improve accuracy using iterations, glmboost optimizes the loss function in a context of generalized linear models (GLMs), fitting variable by variable, which facilitates greater interpretability of the model's internal fits. This methodology not only allows for a detailed analysis of the contribution of each explanatory variable, but also improves the accuracy of predictions in a more transparent manner, contrasting with the visibility limitations of internal adjustments characteristic of algorithms considered "black box" (Hothorn et al. 2023; Saupin, 2022).

Among the main advantages of the glmboost algorithm, its ability to offer robust predictive models without requiring high computational performance stands out, in addition to having intuitive implementation. This is particularly useful in fields that demand a high degree of model explainability, such as in scientific and research applications. However, it is important to recognize that, like other modeling techniques, glmboost is not without challenges, such as the need to manage

multicollinearity and heteroscedasticity in explanatory variables. However, its component-by-component model-focused approach offers a more transparent avenue for understanding and interpreting model fits, mitigating some of the common concerns associated with traditional machine learning methods (Hothorn et al. 2023; Sosnovshchenko et al., 2018).

NAIVE BAYES

Developed based on the theorem created by Thomas Bayes (1702 – 1761), the "naïve Bayes" is a classifier that generates a table of probabilities for the phenomenon in question from the following equation: eq (2).

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)} \quad (2)$$

where, P: is the probability; P(A|B): is the probability of event B given event A; P(A): is the probability of event A; P(B) is the probability of event B.

This algorithm has been used for classification in various scenarios, such as the classification of emails as "SPAM", product recommendation, diagnostics in the health area, and even to control autonomous vehicles (Wickramasinghe and Kalutarage, 2020).

Among the advantages of this algorithm are its processing speed, good prediction for problems that have more than two possible categories, and its ability to predict when the predictor variables are categorical. Generally, because it is understood that the predictor variables are not related to each other, this can be considered a disadvantage since in real problems this rarely occurs. In addition, when a certain category is not in the training database, the algorithm falls on a problem known as a "zero-frequency problem" (Bruce et al., 2023).

ADJUSTMENT, EVALUATION AND CHOICE OF THE MODEL

To adjust the models, the original database was divided into two parts, leaving 80% of the observations for training and 20% for testing.

For the performance evaluation of classification models, five indicators derived from the confounding matrix are commonly used. They are: Accuracy, sensitivity, specificity, precision, and AUC.

The confusion matrix is a data visualization tool that displays how many predictions were made correctly and incorrectly by the model against the original data.

Table 1 shows a generic confusion matrix and the following information can be seen: True Negatives [VN] – number of observations with the actual and predicted value equal to 0. True Positive [PV] – number of observations with the actual and predicted value equal to 1. False Positive

[FP] – number of observations where the actual value is 0 and the predicted value was 1. False Negatives [FN] - number of observations where the actual value is 1 and the predicted value was 0.

Table 1. Generic confusion matrix

Predicted	Results	
	Reference 0	Reference 1
Predicted 0	VN	FN
Foretold 1	FP	VP

Source: Original survey data

It is important to say that the values displayed in the confusion matrix change depending on the cutoff point chosen by the researcher. As the classification models predict a value between 0 and 1, it is necessary for the researcher to determine up to which value, or cut-off point, the prediction should be considered as "non-event", or 0, so that above that it is considered as "event", or 1. For each of the models estimated in this study, the accuracy for all cut-off models was calculated between 0.10 and 0.90. The value chosen for each model was the one that presented the highest accuracy of the predictions in the test database.

Accuracy is the percentage of correct answers of the model in relation to the total of predictions. This result is obtained with the use of eq. (3) and returns a value between 0 and 1.

$$\text{Acurácia} = \frac{VP + VN}{VP + VN + FP + FN} \quad (3)$$

The value obtained represents the percentage of correct answers of the model in relation to the total number of cases. For example, if the result was 0.8456, it is known that the model correctly predicted 84.56% of the total cases.

Sensitivity is the proportion of correct answers in the model in relation to the total number of positive cases. The result is obtained through eq. 4 and returns a value between 0 and 1.

$$\text{"Sensitivity"} = \frac{VP}{VP + FN} \quad (4)$$

A result of 0.7615 indicates that the model correctly predicted 76.15% of the positive cases present in the sample or, in other words, the model predicted 23.85% of the positive cases as negative.

Specificity is defined as the proportion of correct answers in the model in relation to the total number of negative cases. This result is obtained using the eq. (5) and returns a value between 0 and 1.

$$\text{"Specificity"} = \frac{VN}{VN + FP} \quad (5)$$

Similar to the sensitivity, if the result obtained was 0.9150, it can be said that the model correctly predicted 91.50% of the negative cases present in the sample or, in other words, the model predicted 8.50% of the negative cases as positive.

Accuracy is characterized by the proportion of correctly positive predictions within those predicted as positive by the model. The result is obtained using eq. (6) and returns a value between 0 and 1.

$$\text{"Precision"} = \frac{VP}{VP + FP} \quad (6)$$

If the result is 0.9, it can be concluded that 90% of the values predicted as positive were in fact positive or, in other words, 10% of what the model predicted as positive was negative.

Finally , the AUC, or the "Area under ROC curve", is presented, which, according to Šimundić (2009), is a global metric of diagnostic accuracy. This index, which uses as parameters the sensitivity and specificity obtained from the model's predictions, brings a value between 0 and 1 in which any value lower than 0.5 demonstrates that the created model is not able to categorize the observations at an acceptable level, and then the model can be considered as not usable. Also according to the author, these global metrics exist to compare two or more diagnoses to assess which is the most appropriate, which is one of the challenges of this study.

The "Analytic Hierarchy Process" [AHP] method was used as support for the choice of the most appropriate model. According to Saaty (1977), through the AHP, the important factors for decision-making are organized to facilitate the visualization of their relationships and help the decision-maker to compare the elements more assertively through the display of data in the same magnitude.

As a result of the process, the AHP assigns an overall preference index to each of the alternatives analyzed. The alternative that contains the highest index should be considered as the preferred one according to the criteria provided to the method. However, it should not be understood as an absolute factor for decision-making, but as an additional index to be taken into account.

This study used the R language to load, adjust and analyze the data, as well as to generate predictions of death from heart failure using different binary classification models. Table 2 shows the packages used.

Table 2. List of packages used in this study

Package	Description	Version
readr	Read csv file	2.1.4
jtools	Get details about regression model	2.2.2
dply_r	"if_else" function	1.1.4
Carit	Create confusion matrix	6.0-94
Proc	Visualize the ROC curve	1.18.5
e1071	Estimating Vector Machine and Support	1.7-14
rpart	Estimating Decision Tree	4.1.21
rpart.plot	View Decision Tree	3.1.1
mboost	Estimar modelo “gradient boosting”	2.9-9
naivebayes	Estimating the "naïve Bayes" model	0.9.7
unbalanced	Balancing Database	2.1

Source: Original survey results

RESULTS AND DISCUSSION

During the analysis of the data through their descriptive statistics, a significant difference of order of magnitude was observed between some variables, as well as significant differences between their minimum and maximum values, as shown in Table 3.

Differences in magnitude were observed, for example, between the variables "age" and "platelets". The variable "platelets" can also be used as a reference for analyzing the breadth of the data.

Table 4 shows that the dependent variable "DEATH_EVENT" is unbalanced, as it contains only 32% of records equal to 1. This distribution can hinder the fit of some models, as it receives more information related to a specific category, which can generate bias in the forecasts.

Table 3. Basic statistics of discrete and continuous variables in the database

Variables	min	q1	median	average	q3	Max
age	40,0	51,0	60,0	60,829	70,0	95,0
creatinine_ phosphokinase	23,0	116,5	250,0	581,839	582,0	7861,0
ejection_ fraction	14,0	30,0	38,0	38,084	45,0	80,0
platelets	25100,0	212500,0	262000,0	263358,03	303500,0	850000,0
serum_ creatinine	0,5	0,9	1,1	1,394	1,4	9,4
serum_ sodium	113,0	134,0	137,0	136,625	140,0	148,0
time	4,0	73,0	115,0	130,261	203,0	285,0

Source: Original survey results

Table 4. Frequency of categorical variables in the database.

Variables	0	1
anaemia	170	129
diabetes	174	125
high_blood_pressure	194	105
sex	105	194
smoking	203	96
DEATH_EVENT	203	96

Source: Original survey results

In this section, the accuracies obtained in each of the models and in each division of the database are presented. Because these are five distinct models executed both in the original and balanced bases, it is important that the data from each one are presented individually and then that the analysis and contrast of the information is carried out in order to identify which model seems to be the most appropriate and what are its limitations in the scope of this study.

BINARY LOGISTIC MODEL

After performing the stepwise procedure, it was identified that only the variables "age", "ejection_fraction", "serum_creatinine" and "time" are relevant in the 95% confidence interval, as can be seen in Table 5. The coefficients presented in the table below will be discussed in the next section of this study.

Table 5. Stepwise Procedure Response Table

Variable	Estimators	Standard Error	2.5%	97.5%	With Val.	p
(Intercept)	0.8400	1.1465	-1.4071	3.0872	0.7327	0.4637
age	0.0372	0.0163	0.0052	0.0693	2.2754	0.0229
ejection_fraction	-0.0687	0.0162	-0.1004	-0.0369	-4.2381	0.0000
serum_creatinine	0.7171	0.1898	0.3450	1.0891	3.7779	0.0002
time	-0.0205	0.0031	-0.0266	-0.0144	-6.5851	0.0000

Source: Original survey results

Table 6 shows that in the original training database, which contains 80% of the observations, the model achieved an accuracy of 82.4%. On the basis of balanced training, it is possible to observe that accuracy was increased by 2.1%, reaching 84.6%

Table 6. Binary logistic model training confusion matrix

Predicted	Results			
	Original		Balanced	
	Reference 0	Reference 1	Reference 0	Reference 1
	----------%----------			
Predicted 0	143	15	150	12
Foretold 1	27	54	38	124

Source: Original survey results
Note: In the Original base the cutoff point was 0.55 and in the Balanced base it was 0.65

The test database contained 20% of the observations from the original database. The prediction of the model estimated from the original database had an accuracy of 91.7%. The prediction of the model estimated from the balanced base had an accuracy of 90%, a fact that can be seen in Table 7.

Table 7. Binary logistic model test confusion matrix

Predicted	Results			
	Original		Balanced	
	Reference 0	Reference 1	Reference 0	Reference 1
	----------%----------			
Predicted 0	43	2	41	4
Foretold 1	3	12	2	13

Source: Original research resultsNote: In the Original base, the cutoff point was 0.55 and in the Balanced base, it was 0.65

VECTOR MACHINE AND STAND

Even though this algorithm was able to cross-validate the data, two scenarios were executed.

In the evaluation of the model, the cross-validation technique with 5 subsets was adopted, ensuring that each segment of the dataset is used for testing once, with the remaining segments employed in training. The hyperparameters established for the model included a polynomial kernel, a regularization value (C) of 3, and gamma set to 0.1. In addition, it was decided not to standardize the data ("scale" defined as "FALSE") because this strategy did not show any improvement in the performance of the model. With these configurations, an accuracy of 44.8% was observed using the data in its original form, which increased to 47.4% after the process of balancing the classes, evidencing an improvement in the performance of the model with balanced data.

A second attempt was made in this same scenario in which the hyperparameter "scale" was changed to "TRUE". With this parameter, the algorithm standardizes the values of the variables using the "Z-Score" transformation and the accuracy verified for the model with the original data was 91.3% and 93.2% with the balanced data, as can be seen based on Table 8.

Table 8. Vector Machine Confusion Matrix and Bracket

Predicted	Results			
	Original		Balanced	
	Reference 0	Reference 1	Reference 0	Reference 1
	----------%----------			
Predicted 0	201	24	183	17
Foretold 1	2	72	9	175

Source: Original research resultsNote: In the Original base, the cutoff point was 0.45 and in the Balanced base, it was 0.57

In the second scenario, the database was separated into training and testing. Using the same parameters as in the second attempt, the model obtained an accuracy of 91.9% on the original training basis. As can be seen in Table 9, the model with the balanced training base was 95.1%.

Table 9. Vector Machine Training Confusion Matrix and Bracket

Predicted	Results			
	Original		Balanced	
	Reference 0	Reference 1	Reference 0	Reference 1
	----------%----------			
Predicted 0	156	15	155	9
Foretold 1	2	66	7	153

Source: Original research resultsNote: In the Original base, the cutoff point was 0.45 and in the Balanced base, it was 0.57

The test database contained 20% of the observations from the original database. The prediction of the model estimated from the original database had an accuracy of 83.3%. The prediction of the model estimated from the balanced base had an accuracy of 85% according to the confounding matrix in Table 10.

Table 10. Vector Machine Testing Confusion Matrix and Bracket

Predicted	Results			
	Original		Balanced	
	Reference 0	Reference 1	Reference 0	Reference 1
	----------%----------			
Predicted 0	41	6	38	2
Foretold 1	4	9	7	13

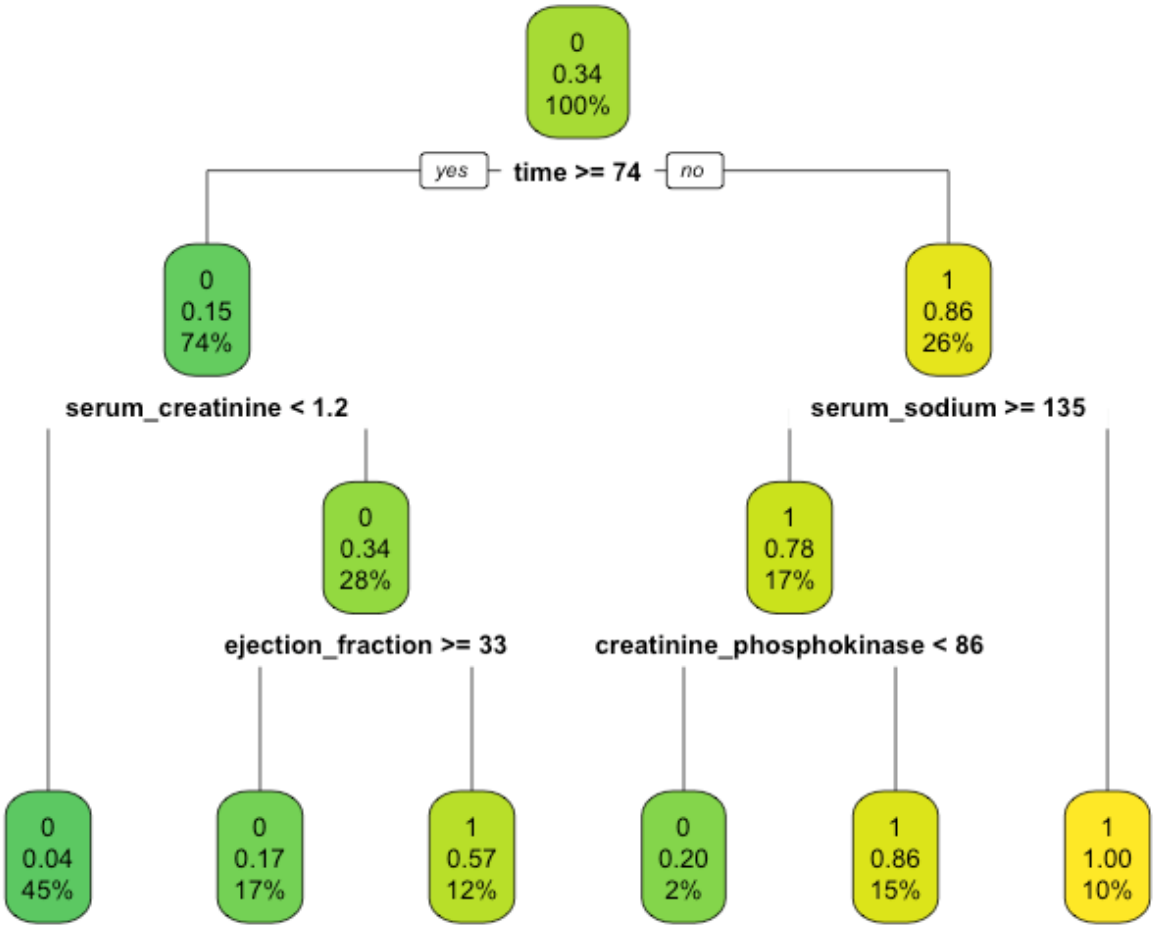
Source: Original research resultsNote: In the Original base, the cutoff point was 0.45 and in the Balanced base, it was 0.57

DECISION TREE

For the execution of this algorithm, the original database was used with all the predictor variables. In Figure 2 it is possible to identify that many branches are not needed for the tree to reach the result. In addition, it is possible to notice that the variables "team", "serum_creatinine",

"serum_sodium", "creatinine_phosphokinase" and "ejection_fraction" were considered to arrive at the result.

Figure 2. Model Decision Tree



Source: Original survey results

The training database, which contained 80% of the data from the original database, had an accuracy of 86.2%. On the other hand, the model with a balanced training base showed an accuracy of 89.8%, according to the data shown in Table 11.

Table 11. Decision tree training confusion matrix

Predicted	Results			
	Original		Balanced	
	Reference 0	Reference 1	Reference 0	Reference 1
	----------%----------			
Predicted 0	153	28	151	22
Foretold 1	5	53	11	140

Source: Original survey resultsNote: The cutoff point was 0.60 in both databases

The test database contained 20% of the observations from the original database. The prediction of the model estimated from the original database had an accuracy of 83.3%. The

prediction of the model estimated from the balanced base had an accuracy of 81.7%, as shown in Table 12.

Table 12. Decision tree test confusion matrix

Predicted	Results			
	Original		Balanced	
	Reference 0	Reference 1	Reference 0	Reference 1
	----------%----------			
Predicted 0	41	6	39	5
Foretold 1	4	9	6	10

Source: Original survey resultsNote: The cutoff point was 0.60 in both databases

GRADIENT BOOSTING

For this algorithm, two models were created to evaluate which would be more effective in predicting death. The first model was performed taking into account only the variables that the stepwise procedure of the logistic regression algorithm indicated as relevant, i.e., with the variables "age", "ejection_fraction", "serum_creatinine" and "time". While it is possible to use the selection of variables from one model to inform another, it is important to note that the assumptions and characteristics of each model are different. Therefore, the variables important for a logistic regression model may not be the most appropriate for a gradient boosting model. The second model was created using all the variables present in the database, but it only considered the variables presented in Table 13 as relevant.

Table 13. Model variables and their respective coefficients.

Variable	Value
(Intercept)	3.939965e+00
age	1.065228E-02
creatinine_phosphokinase	4.992803e-05
ejection_fraction	-2.332532e-02
serum_creatinine	2.348574E-01
serum_sodium	-2.277677e-02
time	-8.056760e-03

Source: Original survey results

The model estimated from the original database obtained the highest accuracy, being 82.8% in the training division. On the other hand, the model estimated from the balanced database showed slightly lower accuracy than the unbalanced model, being 82.1%, as can be seen in Table 14.

Table 14. Gradient boosting training confusion matrix

Predicted	Results			
	Original		Balanced	
	Reference 0	Reference 1	Reference 0	Reference 1
	----------%----------			
Predicted 0	142	16	151	11
Foretold 1	25	56	47	115

Source: Original research resultsNote: In the Original base, the cutoff point was 0.51 and in the Balanced base, it was 0.34

As can be seen in Table 15, the prediction of the unbalanced model presented an accuracy of 90%, while the estimate of the balanced model presented a small increase of 91.6%.

Table 15. Gradient boosting test confusion matrix

Predicted	Results			
	Original		Balanced	
	Reference 0	Reference 1	Reference 0	Reference 1
	----------%----------			
Predicted 0	42	3	42	3
Foretold 1	3	12	2	13

Source: Original research resultsNote: In the Original base, the cutoff point was 0.51 and in the Balanced base, it was 0.34

NAIVE BAYES

As this model does not consider any correlation between the variables, all of them were used for its creation. Table 16 shows that in the original training database, which contains 80% of the observations, the model achieved an accuracy of 26.8%. On the basis of balanced training, it is possible to verify that the accuracy was increased by 3.2%, reaching 29.9%.

Table 16. "Naïve Bayes" training confusion matrix

Predicted	Results			
	Original		Balanced	
	Reference 0	Reference 1	Reference 0	Reference 1
	----------%----------			
Predicted 0	6	23	7	72
Foretold 1	152	58	155	90

Source: Original survey resultsNote: The cutoff point was 0.20 in both databases

The test database contained 20% of the observations from the original database alone. The prediction of the model estimated from the original database had an accuracy of 26.7% and from the balanced database the accuracy was 23.3%, as can be seen in Table 17.

Table 17. Naïve Bayes test confusion matrix

Predicted	Results			
	Original		Balanced	
	Reference 0	Reference 1	Reference 0	Reference 1
	----------%----------			
Predicted 0	4	3	5	6
Foretold 1	41	12	40	9

Source: Original survey resultsNote: The cutoff point was 0.20 in both databases

SUMMARY OF METRICS OBTAINED

To facilitate the visualization and interpretation of the results, Table 18 was created, which, in addition to showing the accuracy achieved, also shows the sensitivity, precision, and AUC of each model in each of the databases.

Table 18. Summary of key metrics from models
(continues)

Model	Results										
		Original					Balanced				
	Divi	PC	Abas	Meaning	Preci	AUC	PC	Abas	Meaning	Preci	AUC
		----------%----------									
Binary logistics	TR	0,55	82,4	78,3	66,7	88,8	0,65	84,6	91,2	76,5	93,1
	TS		91,7	85,7	80,0	87,8		90,0	76,5	86,7	88,9
Gradient boosting	TR	0,51	82,8	77,7	69,1	79,5	0,34	82,1	91,2	70,9	82,7
	TS		90,0	80,0	80,0	86,6		91,6	81,2	86,6	90,0

Table 18. Summary of key metrics from models
(conclusion)

Model	Results										
		Original					Balanced				
	Divi	PC	Abas	Meaning	Preci	AUC	PC	Abas	Meaning	Preci	AUC
	----------%----------										
Vector Machine & Stand	CM	0,45	91,3	75,0	97,3	87,0	0,57	93,2	91,1	95,1	93,2
	TR		92,9	81,5	97,1	90,1		95,1	94,4	95,6	95,1
	TS		83,3	60,0	69,2	75,6		85,0	86,7	65,0	85,6
Decision tree	TR	0,60	86,2	65,4	91,4	81,1	0,60	89,8	86,4	92,7	89,8
	TS		83,3	60,0	69,2	75,6		81,7	66,7	62,5	76,7
Naive Bayes	TR	0,20	26,8	71,6	27,6	37,7	0,20	29,9	55,6	36,7	29,9
	TS		26,7	80,0	22,6	61,1		23,3	60,0	18,4	35,6

Source: Original survey results
Note: [Divi] Division; Training [RT]; Test [TS]; Complete [CM]; Cut-off point [PC]; Percentage accuracy [Acur]; Sensitivity in percentage [Sens]; Precision in percentage [Preci]; AUC in percentage

ANALYSIS OF METRICS

This study considered in its evaluation one of the most popular metrics for binary classification models, accuracy. However, a limitation of this metric is that it considers only a cut-off point, so the researcher must pay attention to it because his choice can significantly change the result of the study. According to Šimundić (2009), accuracy needs to be analyzed with caution, as it only shows that the model is capable of bringing the correct classifications, but concludes by saying that other diagnostic measures should be weighted so that the conclusion is more assertive.

Because it is related to the health area and having as a reference both the popular saying that says that prevention is better than cure and scientific knowledge that recognizes the value of early diagnoses, the second metric chosen was sensitivity. The higher its value, the lower the number of false negatives, that is, the smaller the number of people who should be accompanied by a health professional, but the model was not able to predict.

The third index used was the AUC, because, as previously mentioned, it is a metric that serves to compare different models and, for this reason, it becomes a fundamental piece for the analysis of the results produced by this study.

Another aspect that was weighted was the level of performance gain of the forecasts obtained from the estimated models after balancing the data. Although the technique used in this step has solid theoretical foundations, the observations generated are synthetic. Because it is an analysis of health factors, it was defined that, if the metrics of the models estimated from the balanced database were slightly higher than the metrics obtained from the models estimated from the original database, for reliability reasons, preference would be given to the model estimated from the original data because it does not contain synthetic data.

After defining the metrics and aspects that would be used for the evaluation, the results obtained in the test database of each of the models were analyzed.

The "naïve Bayes" model, because it obtained results that showed the lowest accuracy, precision and AUC, proved to be the least adequate for predicting the response variable.

The decision tree metrics did not have much variation in the test base with both the model estimated from the original base and the model generated from the balanced base. The accuracy achieved was 83.3% on the original basis and 81.7% on the balanced basis. Thus, it can be concluded that, although the balancing has contributed to the generalization of this algorithm, this model was not able to achieve the best accuracy among the models created by this study.

As the vector and support machine algorithm has the ability to "cross-validation", it may be strange that it was estimated both with the complete database and with the division between training and tests. However, it is known that in SVM models "overfitting" *can occur* due to the choice of

hyperparameters. The comparison between the results obtained with the two strategies can be beneficial for a better understanding of the results and analysis of possible overfitting.

The accuracy achieved with the complete balanced database was 93.2% with a sensitivity of 91.1%, precision of 95.1% and AUC of 93.2%. In absolute numbers, this is the best model created in this study. The same model with the unbalanced base obtained an accuracy of 91.3% with a sensitivity of 75%, precision of 97.3% and AUC of 87%.

By implementing the technique of dividing training and testing using the same hyperparameters as the previously estimated models, the model's prediction capacity is greatly affected and the accuracy and sensitivity metrics are close to those achieved in the decision tree.

Based on these data, it is reasonable to assume that the full-base models are, in some way, incurring in "overfitting" since, even using the same parameters, when the generated model receives unknown data, the accuracy is 9.6%, the sensitivity is 21.5% and the precision is 27.8% lower than that achieved when the data are already known. This difference may mean that the SVM models proposed here, although they bring results that can be considered satisfactory, were not sufficiently generic to be considered when used with data unknown to the model.

The gradient boosting model, on the other hand, was able to predict 90% of the cases correctly on the original basis, being the second highest accuracy and also the second highest AUC because we did not consider the full-base SVM model. On a balanced basis, the model predicted 91.6% of the cases correctly. These models also showed high sensitivity and precision, reaching satisfactory results and very close to those achieved in the binary logistic model.

The binary logistic model had its best performance when estimated with the original database with an accuracy of 91.7%, sensitivity of 85.7%, precision of 80% and AUC of 87.8%. When comparing the results of the predictions in the databases, it can be seen that in the original database the accuracy was 1.7% and the sensitivity 9.2% higher compared to the balanced base, however, the precision is 6.7% lower.

It is important to point out that the binary logistic model estimated with the original database has higher accuracy and sensitivity in the test base compared to that achieved in the training base. This may indicate that the model is satisfactorily generic enough to adequately predict the category of an observation that has data not present in the sample at the time it was estimated.

Reinforcing this perspective, the application of the AHP method for analysis of the results consolidates the robustness of the binary logistic model in the original database, since it, although very close to the gradient boosting in the balanced base, receives the highest global preference index. It is worth remembering that one of the main premises of this study was to give preference to the models estimated from the original data and, therefore, when we compare the two best options using

only these, the global preference index of the original binary logistic model is comfortably distant from the gradient boosting estimated from the original base, as can be seen in Table 19.

Table 19. Overall preference index obtained through the AHP method

Model	Global Preference Index
Original Binary Logistic	12,03%
“Gradient boosting” balanceada	11,99%
“Gradient boosting” original	11,70%
Balanced Binary Logistics	11,70%
Vector machine and balanced support	11,39%
Balanced Decision Tree	10,34%
Vector machine and original support	10,30%
Original Decision Tree	10,30%
“Naive Bayes” original	5,83%
“Naive Bayes” balanceada	4,44%

Source: Original survey results

The fact that the binary logistic model is shown to be the most appropriate for this type of prediction is supported in the literature when Fávero and Belfiore (2017) discuss the study of the possibility of infarction as a variable to be predicted and physical and behavioral characteristics as predictor variables. The objective is to estimate the probability of occurrence of this dichotomous event through binary logistic regression.

The AUC of 87.8% achieved with the binary logistic model can be considered optimal, close to excellent, for the diagnosis of accuracy (Šimundić, 2009). Understanding whether the 91.7% accuracy is adequate for the health area is something more complex, because due to biological and demographic complexities, it is not possible to determine a value, or a range of values, that satisfies all cases. However, leaving the numbers and reflecting a little, one can reach a conclusion: the more urgent the solution to the phenomenon studied, the lower the accuracy initially accepted.

An example of this is what was seen during the development of vaccines for COVID-19. In the case of CoronaVac, in phase 3 of the tests, the immunizer had an overall efficacy of 62.3%, which initially can be considered low efficacy, but given the dimension of the impacts generated by the disease on the mental health of the population and the economy, the immunizer was approved by the competent bodies (Butantan Institute, 2022).

On the other hand, the object of this study, although relevant, does not have as significant an impact as the aforementioned pandemic. Therefore, it is necessary to compare it with studies that also focused on heart failure.

In a study very similar to this one, Liang and Guo (2023) achieved an accuracy of 78% and AUC of 71.9%, and complement by saying that this number demonstrates a certain superiority when compared to other recently published studies. Mahmud et. al (2023) brings a meta-model for

predicting heart failure that, using the combination of four machine learning algorithms, obtained an accuracy of 87%.

When comparing the values reached here, with an accuracy of 91.7% and an AUC of 87.8%, with other recent studies on the same topic, it is possible to conclude that this study achieved good indices with regard to the metrics extracted, and it is therefore feasible to be considered as a good starting point for more advanced studies.

ANALYSIS OF THE ESTIMATED LOGISTICS MODEL

As stated in the first pages of this study, one of the positive factors of this model is its ease of interpretation. It is therefore necessary to analyze the independent variables selected with a 95% confidence interval and their respective coefficients, as shown in Table 20, in order to identify the impact of each one on the probability of occurrence of the "DEATH_EVENT" event.

Table 20. Summary of key metrics from models

Variable	Estimators	Standard Error
(Intercept)	0.8400	1.1465
Acts	0.0372	0.0163
ejection_fraction	-0.0687	0.0162
serum_creatinine	0.7171	0.1898
team	-0.0205	0.0031

Source: Original survey results

Therefore, it is perceived that "age" or age, "ejection_fraction" or ejection fraction, "serum_creatinine" or serum creatinine, and "time" or time of patient follow-up in days are the relevant variables for the model. The variable "smoking" would only be relevant for a confidence interval lower than 88%, so it was not considered in the model.

Bosch et. al (2019) confirm that aging increases the chance of heart failure when they show that it occurs in only 0.04% among people aged 18 to 44 years, but this number rises to 20.9% for people over 85 years of age. Therefore, the fact that the variable "age" is relevant is supported by the literature.

Lam and Solomon (2021) suggest a new classification of heart failure based on ejection fraction. Therefore, although it is being studied to improve the values, the study shows that ejection fraction is directly related to heart failure.

Serum creatinine is produced by the death of muscle cells in the body that must be filtered by the kidneys, but the elevation in its levels may suggest diseases in the heart, or kidneys, which can lead to heart failure (BHF, 2023). Thus, the literature also supports the relationship between creatinine and heart failure.

Finally, there is the number of days that the patient is being monitored by a health professional. It seems fair to conclude that the higher this number, the lower the risk of death from heart failure, as the patient would be receiving the appropriate treatment.

Therefore, all the variables selected by the model as relevant have studies that confirm their relationship with heart failure.

Before analyzing the coefficients, it is important to remember that the logistic models return results between 0 and 1, in which 0 represents the absence of chance of the event, 1 represents the certainty of the occurrence of the event. Assuming that the result was 0.56, it should be understood that the probability of the event occurring is 56%, according to the model. If this value is above the cutoff point defined by the researcher, the result is then defined as an event, if it is below it is defined as a non-event.

For the logistic model of this study, the result is given from Equation 7 using 0.55 as the cutoff point.

$$P = \frac{1}{1 + \exp(-(0{,}84 + 0{,}0372 * age + - 0{,}0687 * ejection\_fraction + 0{,}7171 * serum\_creatinine + - 0{,}0205 * time))} \quad (7)$$

where, P: is the probability of the event occurring; exp: is the exponential of the negative of the value obtained by the expression in parentheses.

From it, it is concluded that for each year lived, the model assigns a positive coefficient of 0.0372, that is, for each additional year the possibility of death from heart failure has a small increase. For each additional unit of ejection fraction, which has a negative coefficient of 0.0687, the model decreases the possibility of death from heart failure. Following this logic for the other variables, it can be concluded that the parameter that most increases the risk of death is serum creatinine, keeping the other variables constant. The number of days being monitored is the one that least reduces the risk, according to the model.

When Equation 7 is applied to the observations of the test base, an accuracy of 91.7%, sensitivity of 85.7%, precision of 80% and AUC of 87.8% using the cut-off point of 0.55.

FINAL CONSIDERATIONS

It is believed that this study achieved the proposed objective, as it developed a binary logistic model capable of correctly classifying 91.7% of the observations, a result superior to that of similar studies. It is also admitted that it is possible to incorporate it into health systems both in terms of its classification capacity and in relation to the costs associated with its training, storage, availability for access and execution of forecasts so that its result can be added to the patient's clinical history.

Regarding the technical aspects presented, the importance of testing different models for different contexts is emphasized, because in the case of this study, although very sophisticated models were implemented, it was the logistic regression that generated the best results. Together with the need to test different models is the need to choose the metric to compare the results, since this is an important reference for choosing the appropriate model. Finally, the advantage of verifying the descriptive statistics of the data to identify the need for its balancing is mentioned and, if there is, it is recommended to compare the results obtained with the original and balanced data in order to identify if there was a relevant gain in the generalization of the model with the introduction of synthetic data.

Focusing on analytical aspects, it is important that the researcher is aware of the characteristics of each model, such as knowing that a certain model has a tendency to "overfitting", in order to analyze the metrics obtained in the light of this information. Another important point for binary classifications is the attention to the choice of the cutoff point, since this parameter is central to the quantification of the results achieved. Finally, the researcher must be clear about which metrics are most relevant so that, when comparing the results obtained, he or she is able to choose the one that presents the best balance between the metrics taking into account the objective of the study.

For future research, it is recommended to expand the age range of patients, add a sample of data from local health systems, diversify ethnicities, consider behavioral variables such as whether they practice physical exercise, as well as general information such as abrupt weight gain, development of edema, ankle swelling, among others.

ACKNOWLEDGMENTS

In this section, where I must thank those who contributed in a relevant way to this work, I would like to express my gratitude to my object of this study. I thank, from the bottom of my heart, my heart, because without it, literally, nothing would be possible. Also to Professor Erasmo for all his dedication and commitment, because this certainly raised the quality of the study presented.

REFERENCES

1. Arruda, V. L., Machado, L. M. G., Lima, J. C., & Silva, P. R. S. (2022). Tendência da mortalidade por insuficiência cardíaca no Brasil: 1998 a 2019. Disponível em: <https://www.scielosp.org/article/rbepid/2022.v25/E220021/pt/>. Acesso em: 10 out. 2023.

2. Bosch, L., Assmann, P., Grauw, W. J. C., Schalk, B. W. M., & Biermans, M. C. J. (2019). Heart failure in primary care: Prevalence related to age and comorbidity. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6683237/>. Acesso em: 29 fev. 2024.

3. Boswell, D. (2002). *Introduction to Support Vector Machine* [Tese de Mestrado]. California Institute of Technology, Pasadena, CA, Estados Unidos.

4. British Heart Foundation [BHF]. (2023). *Global Heart & Circulatory Diseases Factsheet*. Disponível em: <https://www.bhf.org.uk/-/media/files/for-professionals/research/heart-statistics/bhf-cvd-statistics-global-factsheet.pdf>. Acesso em: 10 out. 2023.

5. British Heart Foundation [BHF]. (2023). *The heart-kidney link*. Disponível em: <https://www.bhf.org.uk/informationsupport/heart-matters-magazine/medical/kidney-heart-link>. Acesso em: 29 fev. 2023.

6. Bruce, P. C., Stephens, M. L., Shmueli, G., Anandamurthy, M., & Patel, N. R. (2023). *Machine Learning for Business Analytics*. In P. C. Bruce, M. L. Stephens, G. Shmueli, M. Anandamurthy, & N. R. Patel (Eds.), *Concepts, Techniques, And Applications with JMP PRO* (2ª ed.). O'Reilly. Disponível em: <https://learning.oreilly.com/library/view/machine-learning-for/9781119903833/c08.xhtml#head-2-51>. Acesso em: 05 jan. 2024.

7. Cramer, J. S. (2002). *The Origins of Logistic Regression* [Tese de Mestrado]. University of Amsterdam, Amsterdã, Países Baixos.

8. Elreedy, D., Atiya, A. F., & Kamalov, F. (2022). A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning. Disponível em: <https://doi.org/10.1007/s10994-022-06296-4>. Acesso em: 21 fev. 2024.

9. Escolar, V., Lozano, A., Larburu, N., Kerexeta, J., Álvarez, R., Echebarria, A., & Azcona, A. (2021). Prediction of heart failure decompensations using artificial intelligence and machine learning techniques. *Basurto University Hospital*, Donostia, San Sebastián, Espanha. Disponível em: <https://www.rccardiologia.com/files/rcc_22_29_4_431-440.pdf>. Acesso em: 17 jan. 2024.

10. Fávero, L. P. L., & Belfiore, P. (2017). *Manual de análise de dados: Estatística e modelagem multivariada com Excel®, SPSS® e Stata®* (1ª ed.). Elsevier.

11. Ferreira, D. (2020). Insuficiência cardíaca. *Hospital da Luz*, Lisboa, Portugal. Disponível em: <https://www.hospitaldaluz.pt/pt/dicionario-de-saude/insuficiencia-cardiaca>. Acesso em: 16 fev. 2024.

12. Géron, A. (2022). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow* (3ª ed.). Disponível em: <https://www.oreilly.com/library/view/hands-on-machine-learning/9781098125967/ch06.html>. Acesso em: 5 mar. 2024.

13. Gholami, R., & Fakhari, N. (2017). Support vector machine: Principles, parameters, and applications. In R. Gholami & N. Fakhari (Eds.), *Handbook of Neural Computation* (pp. 515-535). Academic Press.

14. Guerbai, Y., Chibani, Y., & Meraihi, Y. (2022). Techniques for selecting the optimal parameters of one-class support vector machine classifier for reduced samples. Disponível em: <https://www.igi-global.com/article/techniques-for-selecting-the-optimal-parameters-of-one>.

15. Hothorn, T., Bühlmann, P., Kneib, T., Schmid, M., & Hofner, B. (2023). *glmboost* [Documentação do software]. Disponível em: <https://www.rdocumentation.org/packages/mboost/versions/2.9-9/topics/glmboost>. Acesso em: 14 mar. 2024.

16. Kotsiantis, S. B. (2012). Decision trees: A recent overview. Disponível em: <https://link.springer.com/article/10.1007/s10462-011-9272-4>. Acesso em: 05 mar. 2024.

17. Instituto Butantan. (2022). CoronaVac provou sua eficácia contra Covid-19 no estudo clínico mais criterioso, feito com profissionais de saúde durante pico de casos. Disponível em: <https://butantan.gov.br/noticias/coronavac-provou-sua-eficacia-contra-covid-19-no-estudo-clinico-mais-criterioso-feito-com-profissionais-de-saude-durante-pico-de-casos>. Acesso em: 13 fev. 2024.

18. Lam, C., & Solomon, S. (2021). Classification of heart failure according to ejection fraction. Disponível em: <https://doi.org/10.1016/j.jacc.2021.04.070>. Acesso em: 29 fev. 2024.

19. Liang, Y., & Guo, C. (2023). Heart failure disease prediction and stratification with temporal electronic health records data using patient representation. Disponível em: <https://doi.org/10.1016/j.bbe.2022.12.008>. Acesso em: 29 fev. 2024.

20. Mahmud, I., Kabir, M. M., Mridha, M. F., Alfarhood, S., Safran, M., & Che, D. (2023). Cardiac failure forecasting based on clinical data using a lightweight machine learning metamodel. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10417090/>. Acesso em: 29 fev. 2024.

21. Park, T. (2021). Behind Covid-19 vaccine development. *MIT Schwarzman College of Computing*. Disponível em: <https://news.mit.edu/2021/behind-covid-19-vaccine-development-0518>. Acesso em: 10 out. 2023.

22. Rohde, L. E. P., Montera, M. W., Bocchi, E. A., Clausell, N. O., Albuquerque, D. C., Rassi, S., Colafranceschi, A. S., et al. (2028). Diretriz brasileira de insuficiência cardíaca crônica e aguda. *Arquivos Brasileiros de Cardiologia*. Disponível em: <https://doi.org/10.5935/abc.20180190>. Acesso em: 16 fev. 2024.

23. Saaty, T. L. (1977). How to make a decision: The analytic hierarchy process. *European Journal of Operational Research, 48*(1), 9-26. Disponível em: <https://doi.org/10.1016/0377-2217(90)90057-I>. Acesso em: 25 mar. 2024.

24. Saupin, G. (2022). *Practical Gradient Boosting: A deep dive into Gradient Boosting in Python* (1ª ed.). Paris: AFNIL. p. 17.

25. Šimundić, A. M. (2009). Measures of diagnostic accuracy: Basic definitions. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4975285/>. Acesso em: 11 fev. 2024.

26. Sosnovshchenko, A., & Baiev, O. (2018). *Machine Learning with Swift*. In: Sosnovshchenko, A., & Baiev, O. (Eds.), *Artificial Intelligence for iOS* (1ª ed.). Packt Publishing, Birmingham, England. Disponível em: <https://learning.oreilly.com/library/view/machine-learning-

with/9781787121515/697c4c5f-1109-4058-8938-d01482389ce3.xhtml>. Acesso em: 05 jan. 2024.

27. Taulli, T. (2020). *Introdução à Inteligência Artificial: Uma abordagem não técnica* (1ª ed.). São Paulo: Novatec. p. 26.

28. Uddin, K. M. M., Ripa, R., Yeasmin, N., Biswas, N., & Dey, S. K. (2022). Machine learning-based approach to the diagnosis of cardiovascular disease using a combined dataset. *Dhaka International University*, Dhaka, Bangladesh. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2666521223000145>. Acesso em: 15 jan. 2024.

29. World Health Organization [WHO]. (2023). Cardiovascular diseases. Disponível em: <https://www.who.int/health-topics/cardiovascular-diseases/#tab=tab_1>. Acesso em: 10 out. 2023.

30. Wickramasinghe, I., & Kalutarage, H. (2020). Naive Bayes: Applications, variations and vulnerabilities: A review of literature with code snippets for implementation. Disponível em: <https://doi.org/10.1007/s00500-020-05297-6>. Acesso em: 20 fev. 2024.

ANNEX

Full R Project and Excel support available for download in https://github.com/danielbaldini/TCC_USP_2024_HeartFailureAnalysis