CSIS 3290 – FUNDAMENTALS OF MACHINE LEARNING

***Project Heart Failure Detection***

Student Name: **Ronaldo Aparecido de Oliveira**

Student ID: **300305192**

Section #: **001**

Instructor: **Samuel Otim**

*Douglas College – CSIS Department*

Summer 2020

# **Part 1 - Discovery**

## **1. Introduction**

The Heart Hospital of Canada is a recognized institution on diagnosis and treatment of cardiovascular diseases, assisting more than 50,000 people by year. The hospital has a high-quality care centre with the best specialists dedicated to better serve the patients, and the last generation equipments in a modern building.

Unfortunately, more than 25% of hospital patients die due to heart failure and the correct diagnosis in the early stages is crucial to determine each patient’s perspective of survival. It is estimated that more 17 million people globally die every year and the main factor is myocardial infarctions and heart failures.

## **2. Hypothesis**

The hypothesis to be tested in this work is whether the levels of creatinine phosphokinase, ejection fraction, diabetes, serum creatinine, serum sodium, smoking, high blood pressure and anaemia are associated with higher likelihood of heart failure, a kind of cardiovascular problem.

## **3. Potential Data Sources**

Searching on the Machine Learning Repository website was found three datasets with information that could help to achieve the objectives:

* Heart Disease Dataset from 1988 with 303 instances – associated to classification task (<https://archive.ics.uci.edu/ml/datasets/Heart+Disease>).
* SPECT Heart Dataset from 2001 with 267 instances – associated to classification task (<https://archive.ics.uci.edu/ml/datasets/SPECT+Heart>).
* Heart Failure Clinical Dataset from 2020 with 299 instances – associated to classification and regression tasks (<https://archive.ics.uci.edu/ml/datasets/Heart+failure+clinical+records>).

However, considering the goals and the datasets, probably the third dataset should be used since it has more relevant information regarding to the issue to be solved.

# **Part 2 – Data Preparation**

The dataset selected was the “Heart Failure Clinical Dataset from 2020 with 299 instances” because the features is more related to the objective of the project and contain all necessary variables. The table 1 shows the information contained in the dataset.

TABLE 1 – Dataset Features

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Description | Data Type | Range |
| age | age of the patient in years | Integer |  |
| anaemia | decrease of red blood cells or hemoglobin | Boolean  (integer) | 0: False  1: True |
| creatinine phosphokinase | level of the CPK enzyme in the blood in mcg/L | Integer |  |
| diabetes | if the patient has diabetes | Boolean  (integer) | 0: False  1: True |
| ejection fraction | percentage of blood leaving the heart at each contraction (%) | Integer |  |
| high blood pressure | if the patient has hypertension | Boolean  (integer) | 0: False  1: True |
| platelets | platelets in the blood in kiloplatelets/mL | Float |  |
| serum creatinine | level of creatinine in the blood in mg/dL | Float |  |
| serum sodium | level of sodium in the blood in mEq/L | Float |  |
| sex | woman or man | Integer | 0: Woman  1: Man |
| smoking | if the patient smokes | Boolean  (integer) | 0: False  1: True |
| time | follow-up period in days | Integer |  |
| death event | if the patient died during the follow-up period | Boolean  (integer) | 0: Survived  1: Dead |

As we can see in the table, all features necessary to process are in the dataset. All features are numeric, the dataset do not contain null values, and no data transformation was necessary, including the dummy transformation.

# **Part 3 – Model Planning and Implementation**

Machine learning has been widely applied to try to predict heart failure (Gallagher, 2019), patient survival probability (Al’Aref, 2019), or detection of clinical features (Weng, 2017). This approach is also being applied to diagnosis based on medical imaging records (Tripoliti, 2017). In all cases, many algorithms are applied because there is not a solution for every problem. Chicco and Jurman (Chicco, 2020), for example, used 10 algorithms to try to find the best one.

Following this same idea due to the problem nature, in this project was applied 11 models using Grid Search to find the best parameters for each none. When the best parameters were determined, then each model was processed to find the final score. The table 2 presents the relation of algorithms used, their best parameters, and final scores.

TABLE 2 – Algorithms Applied

|  |  |  |
| --- | --- | --- |
| Algorithm | Best Parameters | Best Score |
| AdaBoost | learning\_rate = 2, n\_estimators = 25 | 0.889 |
| Bagging | max\_features = 0.5, n\_estimators = 100 | 0.833 |
| Decision Tree | criterion = entropy, max\_depth = 2, splitter = best | 0.889 |
| Extreme Gradient Boosting | alpha = 2, gamma = 3, lambda = 2, max\_depth = 6 | 0.922 |
| K-Nearest Neighbors | n\_neighbors = 25, weights = uniform | 0.644 |
| Logistic Regression | C = 0.1, max\_iter = 200, penalty = l2, solver = liblinear | 0.867 |
| Multi-Layer Perceptron | activation = identity, alpha = 0.0001, learning\_rate = constant, max\_iter = 2000, solver = adam | 0.667 |
| Naïve Bayes | var\_smoothing = 1e-09 | 0.833 |
| Random Forest | criterion = entropy, n\_estimators = 7 | 0.856 |
| Stochastic Gradient Descent | alpha = 0.0001, fit\_intercept = True, penalty = l2 | 0.333 |
| TensorFlow (Deep Learning) | Optimizer = ‘adamax’, activation = ‘sigmoid’ (last layer) | 0.767 |

It is important to highlight that all execution, excepting *tensorflow*, were performed using pipeline to make the process more efficient. However, inside the pipeline we use the model obtained after the Grid Search was performed, i.e. first of all, it was defined the parameters dataset to be tested inside the Grid Search, then the Grid Search was executed, and finally, the best model was applied inside the pipeline to get the final score.

As this problem is a kind of classification with only two outcomes, the selection of the technique was direct and logic. Besides that, the dataset is small, and the quality of information is very good. We did not find any outliers, neither null values.

The original hypothesis was to use a set of features to try to predict, with higher likelihood, a heart failure. Using the classification approach, this verification was facilitated because we were able to use all features to test the hypothesis. Moreover, this approach made easier to understand the quality of the prediction as well as how to interpret it. Using some indicators (accuracy, confusion matrix, ROC curve, and feature importance) and charts, we were able to uncover some hidden characteristics of the problem and helped us to propose some insight.

After the processing, the model that presented the best score was Extreme Gradient Boosting (*XGBoost*) (table 2) which its main results are indicated below (table 3 and 4).

TABLE 3 – XGBoost Results

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Confusion Matrix | | | | | **Predicted** | | | | | | |
| 0 | | | 1 | | |
| **Actual** | | 0 | | 59 | | | 1 | | |
| 1 | | 6 | | | 24 | | |
|  | | | | | | | | | | | |
| *Classification Report - XGBoost* | | | | | | | | | | | |
|  | **precision** | | **recall** | | | **f1-score** | | | **support** | | |
| 0 | 0.91 | | 0.98 | | | 0.94 | | | 60 | | |
| 1 | 0.96 | | 0.80 | | | 0.87 | | | 30 | | |
|  | | | | | | | | | | | |
| **accuracy** |  | |  | | | 0.92 | | | 90 | | |
| **macro avg** | 0.93 | | 0.89 | | | 0.91 | | | 90 | | |
| **weighted avg** | 0.93 | | 0.92 | | | 0.92 | | | 90 | | |

TABLE 4 – Feature Importance

|  |  |
| --- | --- |
| Feature | Importance |
| Time | 0.526 |
| Ejection Fraction | 0.140 |
| Serum Creatinine | 0.138 |
| Age | 0.099 |
| Serum Sodium | 0.098 |

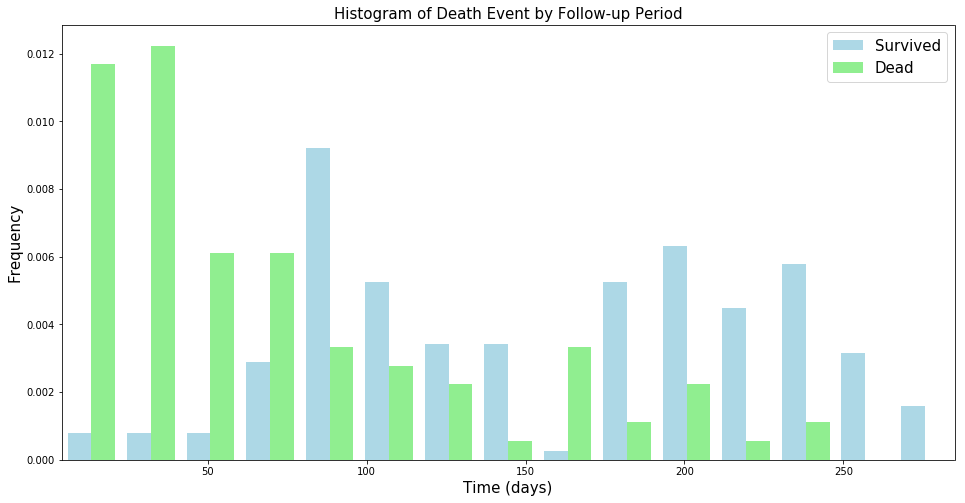
# **Part 4 – Results Interpretation and Business Implications**

Based on the obtained results, we can affirm that:

* The model appears to be valid and very accurate (about 92% - table 3). Its sensitivity (chance of a positive results when the result should be positive) and specificity (chance of a negative results when the result should be negative) are around 91% and 96%, respectively, that are good standards to be applied.
* Considering the experts, the results are coherent. In many resources, experts have been getting similar results, specially Chicco and Jurman (Chicco, 2020) where the precision is very close, and the feature importance differs in 3 variables. Another important characteristic of the model that it is not sensible considering its domain and is not affected by an outlier in the sample (making tolerable to mistake).
* Regards to the goal and the results, we can claim that the model is sufficiently accurate because it can answer our questions with an excellent prediction.
* Although the results are very good, the dataset is very small (only 299 cases), obtained in a short time (April – December 2015) in only one place (Allied Hospital in Faisalabad, Punjab, Pakistan). Thus, a dataset with more patients, from different countries with different eating habits and physical activities, and a longer period of time can improve the quality of the outcomes, making the model more generic. Unfortunately, this dataset does not exist yet.
* Interesting to highlight that the standard form (no customization) of the model was able to address the business problem, just necessary to find the best parameters to get a better score.

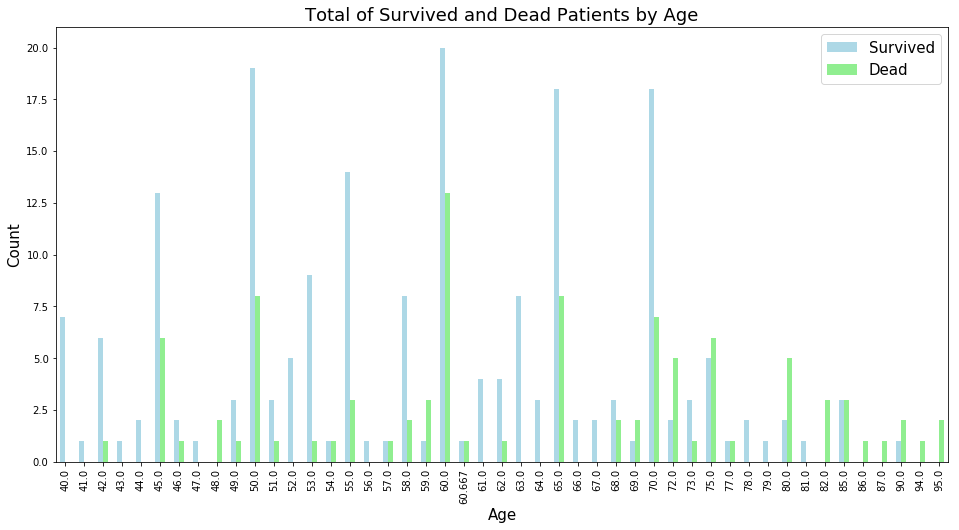
Regarding to major insights, the results showed us that:

1. The 5 most important features to deal with the issue are ‘time’, ‘ejection fraction’, ‘serum creatinine’, ‘age’, and ‘serum sodium’ (table 4).
2. This is quite obvious, but the medical team should try to keep the patient alive as long as possible in the hospital with monitoring to get a health stabilization, helping the recover processing. This follow-up period will provide a better chance of survive after leaving the hospital (figure 1).
3. In the suspicion of a heart failure, it is important to verify the levels of “ejection fraction”, “serum creatinine”, and “serum sodium” (table 4). These values can indicate the severity level of the issue, helping the medical team to understand it, predicting the chance of failure and of the patient surviving in case of failure, besides and better defining the actions require during the follow-up period.



**Figure 1 – Frequency of survived and dead patients by follow-up period**

1. The medical team should pay more attention to patients over 70 years, because after that age, the chance of the patient not surviving is greater than surviving, even with a long follow-up period (figure 2).



**Figure 2 – Total of survived and dead patients by age**

# **Part 5 – Conclusion**

A heart failure is a global problem and kills millions of people every year. To try to find a mode to predict when this issue can occur, it was used a dataset with 299 patients with 12 features related to their medical records. Since it is a classification problem, it was tested 11 algorithms and to help to find the best model, besides the score value evaluation, it was used a tool to perform a set of cross-validation (*GridSearchCV* from *Scikit-Learn*), using a set of predefined values, to determine the best parameters for each model. In the end, the best model was Extreme Gradient Booting (*XGBoost*).

After the processing and analysis, the results showed that the 5 principal features in the model are ‘time’, ‘ejection fraction’, ‘serum creatinine’, ‘age’, and ‘serum sodium’, the longer follow-up period the greater the chance of patient surviving, and the general accuracy of the model is 92%.

Based on the outcomes, the principal business implications are: a) the medical team need to monitor and evaluate the level of “ejection fraction”, “serum creatinine”, and “serum sodium” because this measures can determine the chances of patient surviving; b) also, the team should try to extend the follow-up period and, in this way, minimize the chances of death; c) finally, patients older than 70 years should have greater attention because after this age the risk of death increases more than in other ages.

# **References**

Al’Aref SJ, Singh G, van Rosendael AR, Kolli KK, Ma X, Maliakal G, Pandey M, Lee BC, Wang J, Xu Z, Zhang Y, Min JK, Wong SC, Minutello RM. *Determinants of in-hospital mortality after percutaneous*

*coronary intervention: a machine learning approach*. J Am Heart Assoc. 2019;8(5):011160.

Chicco D, Jurman G. *Machine Learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone*. BMC Medical Informatics and Decision Making. 2020. <https://doi.org/10.1186/s12911-020-1023-5>.

Gallagher J, McCormack D, Zhou S, Ryan F, Watson C, McDonald K, Ledwidge MT. *A systematic review of clinical prediction rules for the diagnosis of chronic heart failure*. Eur Soc Cardiol (ESC) Heart Fail.

2019;6(3):499–508.

Tripoliti EE, Papadopoulos TG, Karanasiou GS, Naka KK, Fotiadis DI. *Heart failure: diagnosis, severity estimation and prediction of adverse events through machine learning techniques*. Comput Struct

Biotechnol J. 2017;15:26–47.

Weng SF, Reps J, Kai J, Garibaldi JM, Qureshi N. *Can machine-learning improve cardiovascular risk prediction using routine clinical data?* PLoS ONE. 2017;12(4):0174944.