ARisCo: Recommendation System for Risk Analysis in Mental Health of Children and Adolescents

Ana Rita Semião e Teixeira

Thesis to obtain the Master of Science Degree in
Information and Software Engineering

Supervisors: Prof. Manuel Fernando Cabido Peres Lopes
Prof.ª Patricia Carla da Silva Pereira

Examination Committee

Chairperson: Prof. Luís Manuel Antunes Veiga
Supervisor: Prof. Manuel Fernando Cabido Peres Lopes
Members of the Committee: Prof. Cláudio Miguel Garcia Loureiro dos Santos Sapateiro

October 2018
Acknowledgments

I would first like to thank my thesis supervisors Manuel Lopes and Patricia Pereira, for the help, support and guidance during all this project. Also, a big thank you to the professors Claudio Sapateiro and Dina Salvador, for all the help and time spent helping me whenever I need it. I would also like to thank all the professionals who were involved in the project: Antonio Nabais, Andre Maravilha, Joana Carvalho, Vera Casimiro, Lusa d’Espiney e Francisca Manso. Without their participation, input and validation this project could not have been successfully conducted.

I would also like to acknowledge the teacher Ana Magalhães as the second reader of this article, and I am grateful to her for the valuable comments and all English corrections. Also professor Isabel Castanheira e Silva for the help provided.

Finally, I must express my gratitude to my family, in particular to my mother, father and sister, to my friends, to my colleagues and last but not the least to my boyfriend and study partner during all these academic years, for providing me with unfailing support and continuous encouragement. This accomplishment would not have been possible without them.

To each and every one of you, thank you.
Abstract

In mental health care, an effective assessment of the risks, is critical to offer good care and safety to the patients. Risks like suicide, self-aggressiveness, heteroaggressiveness and escape must be correctly handled in order to avoid future dangerous situations.

Several factors must be taken into consideration when assigning a mental health patient state, and a correct evaluation of that factors is essential not only to build a solid risk evaluation but also to decide which treatment plan or intervention is the most appropriate.

In this project we propose the development of a new recommendation system to perform risk assessment for the child and adolescent mental health area of hospital Dª. Estefânia. This system aims to facilitate the process of risk assessment currently performed by doctors and nurses manually. It will be responsible to collect relevant information for the risk assignment of the children and adolescents hospitalized and to perform the risk evaluation.

Keywords

Risk Assessment; Recommendation System; Mental Health Care; Data Mining; Decision Trees.
Resumo

No que toca a cuidados em saúde mental, realizar uma avaliação eficaz dos riscos a que os utentes estão expostos, é fundamental para oferecer bons cuidados e segurança aos mesmos. Riscos como suicídio, autoagressividade, heteroagressividade e fuga devem ser tratados corretamente, de maneira a evitar futuras situações perigosas.

Vários fatores devem ser tidos em consideração ao avaliar o estado de um utente de saúde mental e uma avaliação correta desses fatores é essencial não apenas para construir uma boa avaliação do risco, mas também para a escolha de planos de tratamento ou intervenções adequadas.

Neste projeto propomos o desenvolvimento de um novo sistema de recomendação para realizar avaliação de risco para a área de pedopsiquiatria do hospital Dª. Estefânia. Este sistema tem o objetivo de facilitar o processo de gestão de risco atualmente realizado manualmente pelos médicos e enfermeiros. O sistema será responsável por recolher a informação relevante para a avaliação dos riscos das crianças e adolescentes hospitalizados e realizar a avaliação do risco.

Palavras Chave

Avaliação de Risco; Sistema de Recomendação; Saúde Mental; Data Mining; Árvores de Decisão
## Contents

1 Introduction .............................................................. 1
   1.1 Recommendation Systems ..................................... 2
   1.2 Current Situation ............................................... 3
   1.3 Problems and Hypothesis ..................................... 4
   1.4 Project Goal ..................................................... 5
   1.5 Document Structure ........................................... 5

2 Background ............................................................... 7
   2.1 Data Mining ...................................................... 8
   2.2 Classification .................................................. 9
   2.3 Classifier's Process ......................................... 9
   2.4 Metrics of Performance .................................... 10
   2.5 Decision Tree Induction .................................... 11
   2.6 C4.5 Algorithm ............................................... 12
   2.7 Balance of data ............................................... 13
      2.7.1 Synthetic Minority Over-sampling Technique (SMOTE) ... 14
   2.8 Data sample relationships .................................. 15
      2.8.1 Correlation Coefficient ................................. 15
      2.8.2 Binary Logistic Regression ............................. 16

3 Related Work ............................................................ 17
   3.1 Mental Health Care for Children and Adolescents ......... 18
      3.1.1 Risk Management for Children and Adolescents ....... 18
   3.2 Today's Recommendation Systems ........................... 19
   3.3 Health Recommendation Systems ............................ 20
      3.3.1 Health Recommendation Systems in Mental Care .... 20
   3.4 Índice de Risco de Suicídio (IRIS) .......................... 21
   3.5 Galatean Risk and Safety Tool (GRiST) ....................... 21
      3.5.1 Risk Assessment in GRiST .............................. 23
3.6 Risk Assessment in the current instrument ........................................... 24

4 Solution ........................................................................................................ 27

4.1 Data Sample ............................................................................................... 28
    4.1.1 Data Sample’s Demographic Information ........................................... 28
    4.1.2 Current Instrument Evaluation .............................................................. 30
    4.1.3 Risks’ Correlation ................................................................................ 33
    4.1.4 Questionnaire Questions’ Relations ....................................................... 33
    4.1.5 Questionnaire Questions and Risk’s Relations ....................................... 34
        4.1.5.A Correlation coefficient ................................................................ 34
        4.1.5.B Logistic regression \(^1\) ................................................................ 35
        4.1.5.C Main conclusions ......................................................................... 36

4.2 Data Treatment ........................................................................................... 37
    4.2.1 Final Risks’ Levels ............................................................................. 37
    4.2.2 Balancing Data ................................................................................... 37

4.3 Decision Trees ............................................................................................. 38
    4.3.1 Decision Trees Classifiers’ Development ............................................. 38
    4.3.2 Risk category’s counters ..................................................................... 38
        4.3.2.A Pruning techniques performed ..................................................... 39
        4.3.2.B Input questions .......................................................................... 40
        4.3.2.C Cross Validation ..................................................................... 40
        4.3.2.D Health professionals evaluation ............................................... 40
    4.3.3 Final Decision Tree’s classifiers ............................................................. 41
        4.3.3.A Suicide Risk Tree ................................................................... 41
        4.3.3.B Self-Aggressiveness Risk Tree ................................................... 41
        4.3.3.C Heteroaggressiveness Risk Tree ................................................ 42
        4.3.3.D Escape Risk Tree .................................................................... 43

4.4 The new Questionnaire .............................................................................. 43
    4.4.1 Questions’ Removal .......................................................................... 44
    4.4.2 Questions Divided ............................................................................. 45
    4.4.3 Duplicate Questions .......................................................................... 46
    4.4.4 Questions reformulated .................................................................... 46
    4.4.5 New questions and New risk ............................................................... 47
    4.4.6 Final Questionnaire .......................................................................... 47

4.5 Arisco System .............................................................................................. 48

\(^1\)Study performed in collaboration with Professor Dina Salvador, School of Technology, Polytechnic Institute of Setúbal
<table>
<thead>
<tr>
<th>B</th>
<th>Tables</th>
<th>86</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>Code</td>
<td>89</td>
</tr>
</tbody>
</table>
List of Figures

2.1 Knowledge Discovery from Data (KDD)'s process steps \cite{1,2} ......................... 8
2.2 Inputs and outputs of a classifier model ......................................................... 9
2.3 Example of a decision tree for Example 1 ....................................................... 12
2.4 Example of a Logistic Regression (LR) predictor model, relationship between the vari-
   ables: Salary and Job Category \cite{3} .............................................................. 16
3.1 IRIS questionnaire, English version \cite{4} .......................................................... 22
3.2 Information categories collected by GRiST ....................................................... 23
3.3 Input and Output variables of GRiST \cite{5} ....................................................... 24
3.4 Hospital's current questionnaire English version ................................................. 26
4.1 Final decision tree for the Risk of Suicide ....................................................... 41
4.2 Final decision tree for the Risk of Self-Aggressiveness .................................... 42
4.3 Final decision tree for the Risk of Heteroaggressiveness .................................. 42
4.4 Final decision tree for the Risk of Escape ....................................................... 43
4.5 Final questionnaire English version ................................................................. 48
4.6 System Init Page .................................................................................................. 49
4.7 System's architecture of functionality Perform Risk Evaluation ......................... 50
4.8 ARisCo system resume page .............................................................................. 50
4.9 System's architecture of functionality Change fixed question ............................ 51
4.10 System's architecture of functionality Check last evaluations .......................... 52
4.11 Database structure ............................................................................................ 52
5.1 Suicide Tree Classification .................................................................................. 57
5.2 Self-aggressiveness Tree Classification ................................................................ 58
5.3 Heteroaggressiveness Tree Classification .......................................................... 59
5.4 Escape Tree Classification .................................................................................. 59
5.5 Suicide Tree Pilot Test ....................................................................................... 67
5.6 Heteroaggressiveness Tree Pilot Test ........................................... 68
5.7 Escape Tree Pilot Test ............................................................... 69

A.1 Final questionnaire Portuguese version ....................................... 81
A.2 IRIS Portuguese version ............................................................... 82
A.3 Hospital's current questionnaire Portuguese version ...................... 83
A.4 Recommendation System for Risk Analysis in Mental Health of Children and Adolescents (ARisCo) system questionnaire's page .................................................. 84
A.5 Arisco System Poster ................................................................. 85
List of Tables

2.1 Confusion Matrix Template for 2-class [2] .................................................. 10
2.2 Possible Confusion Matrix for Example 1. ..................................................... 10
2.3 Rule of Tumb for interpreting the strength of the correlation coefficient [6] .... 15

3.1 Current hospital’s instrument rules ................................................................. 25

4.1 Data sample’s number of instances per risk category ...................................... 29
4.2 Precision of the current questionnaire rules .................................................... 30
4.3 Sensitivity class Very High of the current questionnaire rules ....................... 31
4.4 Questionnaire’s Rules Confusion Matrix for Suicide Risk .............................. 31
4.5 Questionnaire’s Rules Confusion Matrix for Self-aggressiveness Risk .............. 31
4.6 Questionnaire’s Rules Confusion Matrix for Heteroaggressiveness Risk .......... 32
4.7 Questionnaire’s Rules Confusion Matrix for Escape Risk ............................... 32
4.8 Relation regarding Very High evaluations between the risks: Suicide, Self-aggressiveness, Heteroaggressiveness and Escape. .................................................... 33
4.9 Data sample’s risk classification converted to not Very High and Very High .... 37
4.10 Questions being used on the final decision trees ........................................... 45

5.1 Resulting metrics from the decision trees classifiers versus rules from hospital’s current model ......................................................................................... 56
5.2 Confusion Matrix for Suicide Decision Tree ................................................... 56
5.3 Confusion Matrix for Self-aggressiveness Decision Tree ............................... 57
5.4 Confusion Matrix for Heteroaggressiveness Decision Tree ......................... 57
5.5 Confusion Matrix for Escape Decision Tree .................................................. 58
5.6 Classification performed by the decision trees during the pilot test ............... 60
5.7 Classification performed by the health professionals during the pilot test ....... 61
5.8 Resulting metrics from the decision trees classifiers, Pilot Test .................... 61
5.9 Confusion Matrix for Suicide Risk, Pilot Test .............................................. 61
5.10 Confusion Matrix for Self-aggressiveness Risk, Pilot Test 62
5.11 Performance Metrics for Heteroaggressiveness Risk, Pilot Test 62
5.12 Performance Metrics for Escape Risk, Pilot Test 62
5.13 Confusion Matrix for Risk of Clinic Pathology, Pilot Test 63

B.1 Correlation coefficients between questionnaire’s questions and final risk’s levels - Hospital’s data sample 87
B.2 Correlation coefficients between questionnaire’s questions and final risk’s levels - Pilot Test’s data sample 88
List of Algorithms

C.1 Suicide tree Java Script Object Notation (JSON) ........................................... 90
C.2 Decision Tree Algorithm ................................................................. 91
# Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARisCo</td>
<td>Recommendation System for Risk Analysis in Mental Health of Children and Adolescents</td>
</tr>
<tr>
<td>GRiST</td>
<td>Galatean Risk and Safety Tool</td>
</tr>
<tr>
<td>IRIS</td>
<td>Índice de Risco de Suicídio</td>
</tr>
<tr>
<td>IST</td>
<td>Instituto Superior Técnico</td>
</tr>
<tr>
<td>JSON</td>
<td>Java Script Object Notation</td>
</tr>
<tr>
<td>KDD</td>
<td>Knowledge Discovery from Data</td>
</tr>
<tr>
<td>LR</td>
<td>Logistic Regression</td>
</tr>
<tr>
<td>SMOTE</td>
<td>Synthetic Minority Over-sampling Technique</td>
</tr>
</tbody>
</table>
Introduction

Contents

1.1 Recommendation Systems ........................................... 2
1.2 Current Situation ...................................................... 3
1.3 Problems and Hypothesis ............................................. 4
1.4 Project Goal ............................................................ 5
1.5 Document Structure .................................................. 5
In Portugal psychiatric disorders and mental health problems are one of the main causes of disability and morbidity in the population. The number of children and adolescents who suffer from this mental illness has increased [7]. Although over the last years the attention dedicated to the needs and care of adults and elderly who suffer from mental illness has been increasing, the attention dedicated to children and adolescents who suffer from these diseases is still poor, which may lead to lifelong consequences for these patients [8].

To be able to offer good care and safety to its patients, a hospital must perform a correct management of the available resources, to do that the hospital must have sufficient health professionals and perform adequate treatments and interventions. These treatments and interventions must be done accordingly to the state of the patient [9]. For example, a patient that presents very serious mental symptoms should be frequently observed, to avoid any dangerous situation.

There are not enough nurses to observe all the patients every hour of the day, and possibly not all patients need to be supervised so intensively [8]. Deciding which patients are the ones that need more vigilance is not easy, but the decision should be correctly taken so as to avoid negative scenarios. This decision is made by performing risk assessment, which consists in assessing the likelihood of risk events for the patient and identifying ways for reducing this likelihood [9]. Health professionals must take into consideration the current patient’s symptoms and create a prediction of the state of the patient and future dangerous situations that may occur. Mental health patients that present a very serious risk assessment are the ones that must be observed more intensively. There exist systems that can help in the process of risk management, creating predictions for patients. They are called Recommendation Systems.

1.1 Recommendation Systems

Nowadays systems make predictions all the time to facilitate people’s lives. Information is used to predict future events, which can help the process of decision making. A simple prediction of the weather for tomorrow, may help people plan and decide what to wear or what to do. The same happens in health care: a clinical decision is made based on the analysis of possible outcomes, based on the patients’ symptoms [10].

But not all areas in medicine work equally, each one has its characteristics, risks and treatments, so they cannot be treated as the same [11]. Different areas in medicine require different systems to create these predictions, focused in the characteristics of each one.

Mental health care is different from the rest of medical care in several aspects. An adequate system for mental health care must take into consideration several risks or diseases that other areas do not, because of the different characteristics of people who suffer from mental health problems. For example, it must be taken into account that in this area there exists a probability of the patient committing suicide,
escaping from the hospital, be violent or self-harm [11].

In the same way that mental care has different characteristics from other areas of health care, children and adolescents in mental health care also have different characteristics from adults and elderly [12], and so they need a system adequate for their necessities. A system that makes accurate predictions for grown-up patients, does not necessarily make the same for children or adolescents, because of the factors that must be considered when dealing with younger patients. Evaluating the risks and predicting the possible outcomes for a mental patient is not an easy procedure. An outcome can depend on multiple variables and can vary from patient to patient.

Although it is a complex work, research on mental health care recommendation systems is a work that must be done, because these predictions are essential to the wellbeing of the patients. Between 3.7% - 16.6% of the patients admitted on hospitals in psychiatric areas suffered an adverse event. At least half of them were preventable [11]. An appropriate and correct clinical risk treatment can reduce unnecessary injuries, death and even reduce economic costs for the hospital [11].

1.2 Current Situation

The research on mental health systems is currently not very extensive. However there are several systems that aim to help users in the process of risk assessment, but almost none of them focus in the area of children and adolescents [13].

GRiST, is a decision support system destined to the process of risk management for mental health care that offers a tool focused on the children and adolescents suffering from this illness.

The system is widely used by several hospitals and private institutions. It also can be used by users with no clinical training on the on-line platform [14].

The system consist of a questionnaire used for collecting patients’ information that is relevant to the risk assessment: current state, previous situations, risks the patient are exposed to, family situation, traumatic events and demographic information. After answering all the questions a risk assessment is made and a computerized prediction of the severity state of the patients is created [15].

There exist several tools created for assisting the process of risk assessment for specific risk categories, IRIS [4], is a tool developed to perform an evaluation for the risk of suicide for adult patients which demonstrate having a suicidal ideation of any kind. Although this tool focus only in adult patients and not children and adolescents it offers a good guide to understand what are the most important factors to perform a accurate suicide risk assessment.

Dª. Estefânia hospital is responsible for the treatment of children and adolescents, one of the areas in this hospital is mental health care. The current model used in the hospital for the risk assessment
consists in a list of questions related to five risk categories: suicide, self-aggressiveness, heteroaggressiveness and escape. There are questions regarding each risk category that are filled by the health professionals according to the state of the patient, symptoms that they present and past situations. Based on the answers, a severity level of the risk is predicted for each risk category. Such levels vary according to the scale Very High, High, Medium and Low. The treatment plan, interventions or clinical observation for the patient are decided based on the risk level given by the prediction. It is indispensable that the prediction is correct, so as not to expose children to unnecessary interventions, as well as to both prevent dangerous situations and achieve a better management of the work of the health professionals.

1.3 Problems and Hypothesis

We will improve the risk assessment by creating a model that will be able to perform the same risk predictions as the current one, but using less input variables or different ones. Performing this process automatically will improve the work of the health professionals, so a new recommendation system responsible for processing the risk assessment automatically will be developed. To create the new recommendation system for the hospital, we must understand and correct the failures of the current one. The points of the model that we are going to focus on improving are:

- Identifying which information is relevant for the evaluation of the patient.
  We will use statistical and data mining techniques to find which factors, represented by the questionnaire’s questions, are irrelevant for the risk’s evaluation. We will also try to find correlated factors, the questions concerning one risk indicator that can be used to predict the outcome of another. Besides using these approaches, the health professionals involved in this project will perform focus groups, where doctors and nurses are gathered to discuss which changes should be performed on the current instrument.

- Ambiguous questions, which can lead to inconsistent evaluations depending on the interpretation of each professional.
  This problem can be corrected by reformulating the questions, or/and by splitting one question in several others, creating more specific questions.

- The risk assessment process is made by hand, which unnecessarily consumes time and is conducive to create errors, since the professionals must collect the data and calculate the severity level of the risk manually.
  To improve this point we propose the development of an automatic system that will collect the patient's information and perform the risk assessment process.
1.4 Project Goal

The goal of this project is to develop a new recommendation system for the hospital which will perform the risk assessment process, creating valid and consistent risk evaluations for the patients. The development of this new system will be performed in three stages:

- Analyses: In this first stage the goal is to understand the inconsistencies presented in the current model. To achieve this, data from previous patients of the hospital will be collected and will be used to perform a research using statistical and data mining approaches.

- System Development: After correcting the inconsistencies detected, a new model for risk assessment will be created. Having the new model, the goal is to create a system that will perform this process automatically.

- Validation: Both stages must be submitted to a validation process to guarantee that the model and system are consistent. This validation will be performed using performance metrics and also, by the health care professionals involved in this project to guarantee that the new model and system respect all medical standards. Besides these two types of validation, the new recommendation system will be subjected to a two month pilot test in Dª. Estefânia hospital to guarantee that the new model to perform the risk assessment is accurate and also to guarantee that the system is functioning properly.

The creation of this new recommendation system, responsible for performing risk assessment automatically, is essential for the improvement of the services in mental health care of the hospital not only for the good care of the patients but also to facilitate the work and management of the nurses and doctors.

1.5 Document Structure

Besides this introduction where a summarized overview of all the important points related to the project is made, this document is divided in several other sections:

- Background: This section is constituted by the several concerns that had to be understood to achieve the goals of the project, concerning data mining techniques that were considered relevant and how the current risk assessment tool works.

- Related Work: In this section we describe the current system used for the risk assessment in Dª. Estefânia hospital and related ones: GRIST and IRIS. Besides that, studies made in risk assessment in mental health care area are also described.
• Solution: This section corresponds to the description of the solution of the project, its architecture and development process.

• Evaluation: In this section all the steps related to the evaluation of the solution proposed are described.

• Conclusion: Section dedicated to the conclusions obtained with the project.
## Background

### Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Data Mining</td>
<td>8</td>
</tr>
<tr>
<td>2.2</td>
<td>Classification</td>
<td>9</td>
</tr>
<tr>
<td>2.3</td>
<td>Classifier’s Process</td>
<td>9</td>
</tr>
<tr>
<td>2.4</td>
<td>Metrics of Performance</td>
<td>10</td>
</tr>
<tr>
<td>2.5</td>
<td>Decision Tree Induction</td>
<td>11</td>
</tr>
<tr>
<td>2.6</td>
<td>C4.5 Algorithm</td>
<td>12</td>
</tr>
<tr>
<td>2.7</td>
<td>Balance of data</td>
<td>13</td>
</tr>
<tr>
<td>2.8</td>
<td>Data sample relationships</td>
<td>15</td>
</tr>
</tbody>
</table>
2.1 Data Mining

Data mining is the process of automatically discovering important information from large volumes of data [2]. This information is present in the data by patterns, when discovered and analyzed can be used to predict future outcomes. Data mining is also known as KDD (Knowledge Discovery from Data), which is the process of extracting useful information from data. Although there exist several techniques to perform KDD, they all follow the steps as illustrated in 2.1.

![KDD’s process steps](image)

Data Mining can be divided in two major tasks [2]:

- Predictive tasks: predict a possible value for a attribute based on the current data available.
- Descriptive tasks: find the relations and properties in the data.

Example 1. (Predicting a cancer treatment) [1]

A medical researcher is analyzing data from patients that suffer from breast cancer to predict which one of two treatments is the most appropriate for each one: 'treatment A’ or ‘treatment B’. The data available contains the information about the patient's age and the occurrence of two symptoms: X or Y.

The concept of class is important to understand the data mining’s process. The data entries of a dataset can be associated with a concept or a category, that is called a class [1].

Taking Example 1, the patients’ data contains two classes: ‘treatment A’ class and ‘treatment B’ class because it is possible to categorize the patients into two categories: the ones for whom ‘treatment A’ is appropriate and the ones for whom ‘treatment B’ is appropriate.
2.2 Classification

Classification is a technique of data analysis which belongs to the predictive branch of data mining. It consists of finding the characteristics of several data attributes and assign them to classes [2]. Models are created to perform this process, they are called classifiers and their goal is to predict categorical class labels [1], as illustrated in Figure 2.2. By using classifiers it is possible to predict a class for future entries on the data, based on the categorization learned from the current set of data.

![Figure 2.2: Inputs and outputs of a classifier model](image)

Classifications models are useful to distinguish between data entries from different classes [2]. Taking Example 1, by applying a classifier to the data of patients' information, we can discover which entries (Age, Symptom X and Symptom Y) are associated to the classes 'treatment A' or 'treatment B'. Finding these relations between the data's instances and the classes, it is possible to predict outcomes for future patients. If the classifier learned that people with age \( \leq 30 \) years that suffer from symptom X benefit from treatment A, possibly future patients with these characteristics will also benefit from treatment A.

2.3 Classifier’s Process

The creation of a classifier consists on the following two steps [1]:

- Learning/Training step: where a model is built using a classification algorithm that analyzes the training set and learns the important characteristics and patterns from it.

- Classification step: where a test set is used to estimate the correctness of the model obtained in the previous step.

To perform these two steps the dataset must be divided into two subsets: training set and test set. The training set corresponds to a large subset of the data destined to train the classifier so it can learn its important characteristics. The test set in a smaller subset of the dataset, that must be composed by different instances from the training set, it will be used to evaluate the classifier’s performance created using the training set.

In some cases, the number of instances of the data set is not large enough to be divided into two subsets
and maintain its important characteristics. If the subsets do not have a sufficient number of instances it is impossible to perform training and classification correctly.

One possible solution to this problem is a technique called Cross-Validation. K-fold Cross-Validation is a process that randomly splits the data into k number of subsets with the same size and the two steps, training and testing, are performed k times. In each iteration a random subset is used as the test set and the rest as the train set. After k interactions all the possible combinations of the k subsets have been performed and the average between the results of each iteration is calculated.

## 2.4 Metrics of Performance

The performance of a classifier can be evaluated by comparing the associated class of each entry with the classifier’s class prediction [1], counting the correct and incorrect predictions made by the model. These counts are stored in a table, providing a clear resume of the instances correct or incorrectly classified. This table is known as confusion matrix, and is represented for a 2-class dataset in Table 2.1. The diagonal of the matrix represents the entries that were correctly classified. They are called True Positives (TP) and True Negatives (TN). The rest of the cells correspond to the instances incorrectly classified. The right top cell correspond to the entries that were classified as being from class 1 and actually are from class 0, called False Negatives (FN). The left bottom cell represent the entries classified as class 0 but in reality they belong to class 1, these cases are called False Positives (FP).

<table>
<thead>
<tr>
<th>Actual Class</th>
<th>Predicted Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 0</td>
<td>Class 1</td>
</tr>
<tr>
<td></td>
<td>TP</td>
</tr>
<tr>
<td>Class 1</td>
<td>FP</td>
</tr>
</tbody>
</table>

In Table 2.2 is illustrated a possible confusion matrix obtained from applying a classifier to the data of Example 1. 10 entries were correctly classified as belonging to ‘Treatment A’ (TP) class and 14 to ‘Treatment B’ class (TN). 5 classification errors were committed: 3 instances were classified as being from ‘Treatment A’ class but actually belong to ‘Treatment B’ class (FP), 2 instances were classified as being from ‘Treatment B’ and actually are from ‘Treatment A’ class (FN).

<table>
<thead>
<tr>
<th>Actual Class</th>
<th>Predicted Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment A</td>
<td>10</td>
</tr>
<tr>
<td>Treatment B</td>
<td>3</td>
</tr>
</tbody>
</table>

With the information displayed on the confusion matrix is possible to determinate how well a classifier
performs. But it is important to summarize this information in a single number so that it will be possible to compare the performance of different classifiers [2]. For that performance metrics are used. The metric Accuracy represents the percentage of correct instances classified in all the data; it can be calculated by the formula 2.1.

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}
\]  

(2.1)

The metric of Error Rate represents the percentage of instances incorrectly classified in all the data and can be calculated by the formula 2.2.

\[
\text{Error Rate} = \frac{FP + FN}{TP + TN + FP + FN}
\]  

(2.2)

Sensitivity, also known as True Positive Rate, measures the entries correctly classified as positive; it is calculated using the formula 2.3.

\[
\text{Sensitivity} = \frac{TP}{TP + FN}
\]  

(2.3)

Specificity, also known as True Negative Rate, measures the negative entries correctly identified, calculated through the formula 2.4.

\[
\text{Specificity} = \frac{TN}{FP + TN}
\]  

(2.4)

The metric of Precision represents the exactness of the model, which is the percentage of entries that are classified as positive that actually are; it is calculated with the formula 2.5.

\[
\text{Precision} = \frac{TP}{TP + FP}
\]  

(2.5)

### 2.5 Decision Tree Induction

Decision tree induction is a classification technique consisting on the learning of decision trees from class-labeled training data. A decision tree is a structure based in a tree where each internal node denotes an attribute, each branch node's represents the possible outcomes for that attribute. The topmost node is called root node and the terminal/leaf node holds a class [1].

A tree is constructed by performing tests on the attributes and divide their possible behaviors based on the values they can take [2]. Taking Example 1, for the attribute age, if the two possible values for this attribute are '≤30 years' or '>30 years', the node 'Age' in the decision tree will have two branches, each one corresponding to each possible value. An example of a decision tree representation for Example 1 is illustrated on figure 2.3.

There exist many possibilities of constructing decision trees for a set of data. The accuracy of the
classifier differs from approach to approach and it is impossible to try them all to discover which is the more adequate, which would take too much time. Efficient algorithms for decision trees’ construction are used to create the classifiers with a reasonable accuracy in an acceptable amount of time [2]. The goal of these decision trees algorithms is to find the best attributes that differentiate the entries of a dataset, in order to have clear categories/classes for each entry.

2.6 C4.5 Algorithm

C4.5 is a decision tree algorithm created by Ross Quinlan in the early 1980s [1]. In 2007 it was considered one of the ten best algorithms for treating large amounts of data [16]. It belongs to the algorithms’ family that use a greedy strategy. These algorithms construct the decision tree in a top-down recursive divide-and-conquer approach. The tree is created by analyzing the data training set and it is recursively partitioned into smaller subsets which are represented in the tree [1].

As any other decision tree algorithm, C4.5 receives a set of training data and has the function of choosing the attributes for the tree nodes that best differentiate the data’s entries. At each iteration the algorithm performs the following tasks [17]:

• Creation of a tree node choosing the attribute that better splits the data in that step.

• Choose the branches for the node created on the previous step, corresponding to the possible values for the attribute being represented in the node.

• Create sub-nodes for each branch created.

In the C4.5 algorithm, the nodes that differentiate data most effectively are chosen using the parameter gain ratio. This parameter balances the information gain with the number of possible values for each attribute being tested to become the node. At each iteration, the gain ratio is calculated for all the
attributes, the one that has the higher value is chosen for the tree node. This parameter is calculated using the formula 2.6, being \( p \) a position in the tree and \( \text{attr} \) the attribute being tested to be the node on that position [18]:

\[
\text{GainRatio}(p, \text{attr}) = \frac{\text{Gain}(p, \text{attr})}{\text{SplitInfo}(p, \text{attr})}
\] (2.6)

The numerator of the formula 2.6 stands for information gain, which is the difference in the entropy of the data [18]. The entropy of a dataset corresponds to the purity of the data. If the data only contains entries of one class than the data set is considered to be perfectly pure and the value of its entropy is zero. If the data represents an equal number of entries for each class, the data is said to be impure as possible and has its maximum value, one [2]. The parameter information gain of an attribute tells us how much will be gained if we choose that attribute to be the node of the tree at that point, it can be calculated using the formula 2.7.

\[
\text{Gain}(p, \text{attr}) = \text{Entropy}(p) - \sum_{j=1}^{n} (p_j \times \text{Entropy}(p_j))
\] (2.7)

With \( p \) being a position in the tree, the value of the Entropy for that position can be calculated using the formula 2.8 [18]:

\[
\text{Entropy}(p) = -\sum_{i=1}^{n} (p_i \times \log(p_i))
\] (2.8)

The denominator of the formula 2.6 stands for split information which is the entropy of the dataset concerning only the values of the attribute being tested, it can be calculated using the formula 2.9. The parameter \( P'(j,p) \) represents the number of elements at position \( p \), that take the value of \( j \)-th attribute being tested [18].

\[
\text{SplitInfo}(p, \text{attr}) = -\sum_{j=1}^{n} P'(j,p) \times \log \left( \frac{P'(j,p)}{p} \right)
\] (2.9)

### 2.7 Balance of data

When the data sample has a significant difference in the number of instances of each class, we say that the data sample is unbalanced. The class having the lower number of instances is called minority class and the class having the higher number of instances majority class. An unbalanced data set can bring several problems when developing a predictive model, such as a classifier. The penalty of misclassifying an instance from the minority class can be higher than misclassifying one from the majority class, because the class of interest in the data sample is typically the minority one [19]. For example, in a data sample collected from a medical experiment to perform a study of a way to predict if a person has cancer or not, there would exist a minority class being 'Person with cancer', that
occurs much less frequently than the opposite ‘Person without cancer’, but the class with interest for the classifier is the minority one ‘Person with cancer’ although it occurs less times [1].

Unbalanced data has influence on the performance metrics of the classifiers models, because a classifier by only evaluating correctly the instances from the majority class has a good accuracy, ignoring the instances from the minority class because they are so few that do not influence the global accuracy of the classifier. For example, by developing a classifier model using an unbalanced dataset, having 95% of instances belonging to the positive class and 5% to the negative, the classifier will have a 95% of accuracy by classifying all the instances as being of the positive class, ignoring the minority one.

Sensitivity presented in formula 2.3 and specificity presented in formula 2.4 are metrics that can be used to evaluate the classifier’s performance instead of accuracy. When a classifier presents a very high sensitivity metric (true positive rate), means that the classifier performs well evaluating the positive class. If the metric value for the specificity (true negative rate), is low comparing to the sensitivity the classifier takes more in consideration the correct evaluation of the positive class than the negative one [1], the values of these metrics must be balanced.

There are several methods to address the problem of unbalanced data, the ones being considered in this project are the sampling methods. Sampling methods contains two major techniques to balance the data sample [1, 19]:

- Oversampling: re-sampling the dataset class to contain an equal number of positive and negative instances, by adding new instances to the minority class similar to the existing ones.

- Under-sampling: decreasing the number of instances of the majority class so that the number of the positive and negative class instances in the dataset will stay equal.

2.7.1 SMOTE

SMOTE [20], belongs to the oversampling family of techniques used to balance data. This algorithm creates synthetic instances similar to the ones existing in the data of the minority class [1]. It uses the nearest neighbors of each instance in the data to create synthetic ones for the minority class. The algorithm creates the data following two main steps [20]:

1. It finds the k-nearest-neighbors for minority class instances. By doing this the algorithm has the k instances more similar to the specific one chosen randomly in the minority class.

2. It randomly chooses one of the k-nearest-neighbors and uses it to create a similar one, the new synthetic instance.

Given the number k of nearest neighbors to be used and the percentage of how much new synthetic instances are needed, SMOTE algorithm creates them using the data sample and its characteristics, populating the minority class and balancing the data.
2.8 Data sample relationships

In a data sample it is possible to have relationships between its variables, for example a dataset containing two variables being a person’s age and a person’s weight it is likely that these two variables have a relationship since when a person age increases their weight will normally also increase. There exist statistical techniques to find the hidden relationships in a data sample, in section 2.8.1 the technique of correlation coefficient is described and in section 2.8.2 the LR technique is described.

2.8.1 Correlation Coefficient

Correlation is a statistical technique to determine the relationships between two variables. These relationships are measured by the correlation coefficient, a number that represents the strength of a relationship between the variables. This number can vary between the values -1 and +1 [6]. If the correlation coefficient is close to 0 it means that there are no relationship between the variables, if it is -1 or 1 it means that there exists a strong relationship between them. The level of correlation between the variables can be classified using the rule of Thumb, present in table 2.3.

<table>
<thead>
<tr>
<th>Correlation Coefficient</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.90 to 1.00 (-0.90 to -1.00)</td>
<td>Very strong positive (negative) correlation.</td>
</tr>
<tr>
<td>0.70 to 0.90 (-0.70 to -0.90)</td>
<td>Strong positive (negative) correlation.</td>
</tr>
<tr>
<td>0.50 to 0.70 (-0.50 to 0.70)</td>
<td>Moderate positive (negative) correlation.</td>
</tr>
<tr>
<td>0.30 to 0.50 (-0.30 to -0.50)</td>
<td>Weak positive (negative) correlation.</td>
</tr>
<tr>
<td>0.00 to 0.30 (0.00 to -0.30)</td>
<td>No correlation.</td>
</tr>
</tbody>
</table>

There exist several ways to calculate the value of the correlation coefficient depending on the nature and characteristics of the data. Pearson correlation coefficient is one way to measure the correlation between a pair of variables. For a correlation between the variables \( x \) and \( y \) in a data sample of size \( n \), the formula to calculate the Pearson correlation coefficient is present in 2.10:

\[
\text{PearsonCoefficient}(n) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{2.10}
\]

The formula 2.10 can be simplified, using the covariance and standard deviation. The numerator of the formula is the covariance between two variables, that shows how a pair of variables move in the same direction, positive covariance, or in opposite directions, negative covariance. For example, having two variables being the age and the height of a person, when the age grows the person’s height also grows these variables present a positive covariance. When replacing the numerator with the covariance, the denominator of formula 2.10 is replaced with the standard deviation of the two variables. The standard deviation, represented by the greek letter \( \sigma \), shows the amount of variations presented in the variable’s
values. By making these changes it is possible to reach a simpler formula to calculate the Pearson correlation coefficient
\[\text{PearsonCoefficient} = \frac{\text{cov}(x,y)}{\sigma_x \sigma_y}\] (2.11)

2.8.2 Binary Logistic Regression

Binary LR analysis is one of the mostly preferred regression methods used for analyzing binary or dichotomous dependent variables. A regression analysis has the goal of predicting the dependency of a dependent variable to other independent variables [22]. LR creates a non linear predictor using the same approach of linear regression, this predictor is developed through a graph line defined by an equation created to fit the existing dataset, the relationships between the variables can be categorized through its stronger or weaker relationship, an example of a LR model is illustrated at figure 2.4 where a predictor was created to verify the relationship between the variables salary and the probability a job category [3]. The model illustrated in figure 2.4 has the power to make predictions using the information of variable salary. The model’s predictive line is created using probabilistic approaches that can vary from tool to tool. Observing the graph presented in figure 2.4 is possible to conclude that there are a strong relationship between the variable salary and the probability of the job category. Contrary to what happens in the graph where all the dataset points are positioned in a clear predictive line, if the points are randomly scattered we can conclude that the variables are not related. [3] In figure 2.4 it is also described that there are two sections of weak relationship, in LR the tails of the distributions are the ones where the variable’s relationship is weaker, the section in the middle of the line is where the relationship between the variables and the predicted probabilities is stronger [3].
3

Related Work

Contents

3.1 Mental Health Care for Children and Adolescents .......................... 18
3.2 Today’s Recommendation Systems ........................................... 19
3.3 Health Recommendation Systems ........................................... 20
3.4 IRIS ................................................................. 21
3.5 GRIST ............................................................. 21
3.6 Risk Assessment in the current instrument .............................. 24
To understand the work that must be done, a research was made covering several aspects related to the project scope. First, the research focused on the state of the mental health care concerning children and adolescents, described in section 3.1. A research on recommendation systems was also made, presented in section 3.2 and 3.3. Section 3.5 refers to the risk assessment tool GRiST and section 3.4 to the suicide risk assessment tool IRIS. Finally, for understanding the risk assessment's model currently used in the hospital, in section 3.6 is explained its factors, rules, and evaluation process.

3.1 Mental Health Care for Children and Adolescents

Although a lot of children and adolescents have a normal and happy childhood, about 10% to 20% reveal psychological disturbances. One in five children presents mental problems and only 1/5 of these children receive an appropriate treatment [23].

Some adults that suffer from mental diseases presented symptoms early in childhood or adolescence, which could have been treated at an early stage of the disease in a more efficient way than with a later treatment [24]. Sadly, the area of mental health care focused on children and adolescents does not receive so much attention as it happens with adults and the elderly [8], but it is extremely important, seeing that treating children and adolescents with mental disorders can decrease the number of adults suffering from this kind of illness.

With the increasing number of occurrences of mental health problems in children and adolescents and the necessity of interventions, technical resources are insufficient [24]. For these reasons, the diagnostic of psychological diseases, evaluation of the risks and the choice of the appropriate treatment plan, have become a priority in health care [23].

3.1.1 Risk Management for Children and Adolescents

Such as other diseases, psychological disturbances have clinical criteria used to create a patient’s diagnostic. To produce an efficient diagnostic, a correct identification and evaluation of the patients’ state must be done [24]. The criteria used for the creation of a diagnostic differs from area to area in medicine. Mental health care differs from other areas because the characteristics of the patients and the illnesses are different.

Patients that suffer from mental illness are different, mainly because sometimes they do not realize that they are sick and are hospitalized against their will [11]. The situation is not the same as the one of a patient that is hospitalized with a broken leg and is aware of the necessity of staying in a hospital to get better. In mental health care, issues due to safety of the patient and the health professionals must be taken in account more than in the rest of the medicine areas. Risks like suicide, self-aggressiveness, heteroaggressiveness and escape from the hospital must be taken in consideration and must be early
predicted in order to avoid dangerous future scenarios. These risk predictions are made by identifying the variables most frequently or strongly associated with the risk at stake. Such process is called risk assessment [25]. After performing the risk assessment, an analysis of the resulted prediction is done and it is decided which treatment plan is more adequate to the patient. This process is called risk management. The goal of risk management is to reduce the probability of negative scenarios to occur. Of course even with a correct prediction not all risks are completely eliminated and every treatment plan carries some risk [9]. Risk management is a cycle that is performed several times during the patient’s hospitalization.

Performing risk management to grown-up patients is different from dealing with children and adolescents [12], because the patients’ characteristics change and different factors must be taken in consideration to produce consistent outcomes. When evaluating younger patients, the following risks must be taken in consideration [23]:

- Family dysfunction, such as: early traumatic experiences, situations of violence, mistreatment or high conflict in the family,
- Parents’ behavior,
- Family’s economic difficulties,
- Child’s difficult temperament,
- Situations of vulnerability of the child, specially those that reduce the capacity of control of impulses (epilepsy, brain injury),
- Learning difficulties.

### 3.2 Today’s Recommendation Systems

There exist different definitions for a recommendation system: is a combination of software tools and techniques that provide suggestions for items to a user; is a system that gather several kinds of information to build recommendations [26].

Recommendation systems were created due to the increasing amount of information available on the web with the goal of help users identifying products and services [27]. It is now mostly used in web applications of e-commerce, like Amazon, Netflix or Tripadvisor, to suggest several items that might be of interest to users. On Amazon it can suggests books the user might be interested in buying, on Netflix it suggests movies [28].

These suggestions are presented on a day-by-day basis and aim at helping the users in the decision-making of various subjects: what book to buy, where to eat, which hotel or flight to reserve.
Recommendations made by systems are based on the profile of the user. Systems try to predict which are the most appropriate products or services based on user’s preferences and also the constrains or routines that are collected continuously every time the user uses the system [26]. Taken as an example a user having an account on Amazon. If such user mainly performs researches for cuisine books, the recommendation system will learn that same user is interested in this kind of books and will recommend similar items. Recommendation systems use statistic, machine learning and data mining techniques [28] to predict what items the user should have interested in.

These systems were not only created to help the users searching items. Organizations and companies benefit from them too. The main benefit coming from the use of these systems is the increasing number of items sold. Since the items suggested should suit the user's needs and wants [26], once the item is suggested the user will probably buy it.

3.3 Health Recommendation Systems

Just like information available on the web has been increasing [27], the amount of data collected in clinical databases concerning different kind of patients’ information, has also been increasing drastically [29]. To face this problem, health care has become a new area for the use of recommendation systems. The goal of a recommendation system in medical care is to supply to its user medical information that can be useful to a patient’s medical treatment [29]. Nowadays recommendation systems are trusted and used in the most diverse applications, but in areas concerning health such as diagnoses or treatments, they are still not trustworthy [30].

Recommendation systems destined to the area of health are more complex than the systems used in sales and commerce. These systems are described as being low-risky, that is, if the recommender system of an on-line shop fails calculating which item it should suggest the user to buy, the consequence of this failure is not very serious. In a health recommendation system such a failure can have severe consequences in the patient's state [31].

Health professionals can benefit from these systems, since they can provide additional information to clinical cases, supplying recommendations of likely patients’ diagnoses [31]. These recommendations can help doctors and nurses deciding what the illness in stake is or which is the more adequate treatment for the patient.

3.3.1 Health Recommendation Systems in Mental Care

The number of recommendation systems destined to health care and researches in this area is small [31]. Some models and systems were developed to help in the risk management process, facilitating the evaluation and understanding of the risks patients are exposed to.
3.4 IRIS

IRIS, is a model developed by health professionals from Medical school and Nursing school from Coimbra [4]. This tool was developed to perform a suicide risk assessment for adults patients that present a suicidal ideation. IRIS was developed using a data sample from old patients. Within the 1300 instances, 5% of the patients were selected, the ones that were considered adequate to the study, for example survivors from suicidal attempts.

The tool can perform the risk assessment of an adult patient very simply using a number scale which increases with the answers given to the questionnaire, the english version of the tool is present in figure 3.1, the original tool’s portuguese version is presented at figure A.2. This questionnaire is filled in by the health professional, depending on the answers there is a weight given for each question. These weights are summed and the final weight value indicates with level of suicide risk the patient have.

The weights for assessing the risk are:

- Total weight value $< 5$: Low Risk of Suicide
- Total weight value $\geq 5$ and $< 10$: Medium Risk of Suicide
- Total weight value $\geq 10$: High Risk of Suicide

Observing the IRIS tool presented on figure 3.1, is possible to conclude that some factors have a higher influence on the suicide risk than others. Being an adult male has more weight on the final score than being a female, if the patient is over 45 years old or religious it scores more on the final suicide risk level than if it is not. It is also possible to conclude that having a plan for a suicide attempt is the most important factor when assessing the risk of suicide, this factor adds up 20 points to the final score. Even if the patient do not score in any other questions, only by having this factor is evaluated as a high risk of suicide.

3.5 GRiST

GRiST is a clinical decision support system destined to assist the processes of risk assessment and management in mental health care. It was developed in 2003 by Christopher Buckingham from Aston University in collaboration with researches from Warwick Medical School [32].

The goals which lead to the development of GRiST are [15]:

- Providing universal access to expert advice on risk judgments that can be clearly understood by people without a specialist mental-health background.
- Facilitating risk communication across the care pathway and give to patients more involvement in monitoring and managing risks.
GRIST is a system destined to be used by users without any clinic training and also by clinical experts as an auxiliary tool. The system collects data from users and provides risk estimates [15]. It contains a cloud-computing service, that makes data securely available for consultations whenever needed by the patient or health professional. A patient can provide assessments and get advice from a computerized risk expertise. The doctors can evaluate the assessment data and advise when the patient needs treatment [33]. Detecting risks associated with mental health disorders is difficult for health professionals and even more for users with no clinical knowledge. The idea behind offering knowledge and understanding to users without mental health training is to ask the users the right questions and collect relevant data. Once the data is collected, a risk estimation is created by a computerized expertise, and a risk prediction based on the questions’ answers is provided [33].

<table>
<thead>
<tr>
<th>SOCIODEMOGRAPHY - Weighting 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender: Male → 1  Female → 0</td>
</tr>
<tr>
<td>Age: ≥ 45 → 1  &lt;45 → 0</td>
</tr>
<tr>
<td>Religiosity: No → 1  Yes → 0</td>
</tr>
<tr>
<td>Are there any religious or spiritual factors that could slow down the passage?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CONTEXTS - Weighting 2  No → 0  Yes → 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isolation - do you live alone, without family or social support?</td>
</tr>
<tr>
<td>Significant recent loss - mourning, unemployment, material loss or status</td>
</tr>
<tr>
<td>Physical illness - incapacitating or terminal</td>
</tr>
<tr>
<td>Current abuse of alcohol or substances</td>
</tr>
<tr>
<td>Severe psychiatric illness - current decompensation of psychosis, major unipolar or bipolar major depression, severe personality disorder</td>
</tr>
<tr>
<td>History of psychiatric hospitalization</td>
</tr>
<tr>
<td>Family history of suicide</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SUICIDE SPHERE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personal history of suicidal behavior</td>
</tr>
<tr>
<td>Weighting 3: No → 0  Yes → 3</td>
</tr>
<tr>
<td>Consider Yes in case of 2 or more previous behaviors or only 1 if serious (violent method or having justified intensive care)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Suicidal plan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is the existence of an organized, consistent, lethal, and feasible plan established?</td>
</tr>
<tr>
<td>- valuing recent preparatory acts (ex: farewell letter, will), and access to lethal means (ex: firearm, pesticides / herbicides)</td>
</tr>
<tr>
<td>No → 0  Yes → Directly assign the value 20 to the Total Index Score</td>
</tr>
</tbody>
</table>

TOTAL SCORE ______
Any kind of user can access the GRiST questionnaire on-line and make an evaluation [14], also a paper version of the questionnaire is available. Currently there exists three different versions of GRiST’s questionnaires, each one adapted to different target patients: Working-Age Adults (in this category are included patients between the ages of 18-65 years), Older Adults and Younger People (Children and Adolescents) [15].

3.5.1 Risk Assessment in GRiST

GRiST records user data and provides risk estimates for suicide, self-harm, self-neglect, vulnerability and harm to others [15]. To create these estimations of the risks, a detailed collection of information is done concerning the user’s situation, focused on the factors, current and past life situations that have influence on the risks. This information is collected from the answers given in the GRiST questionnaire, in figure 3.2 the input variables that the system collects are shown.

All the information collected presented at figure 3.2, will indicate the severity of a particular risk of a patient, denominated as risk indicator. The indicators used to evaluate a person’s state may vary in the different approaches of mental health care. In GRiST the risk indicators used are: Suicide, Self Harm,
Harm to Others or Damage, Self Neglect and Vulnerability [15]. There exist several definitions of these risk indicators. Here a brief explanation of each one is presented [13, 34]:

- **Suicide**: Attempt of thought or fantasies of any act with the propose of committing suicide.

- **Self-Aggressiveness**: Attempt of a patient injuring themselves.

- **Harm to Others or Damage**: The patient causes situations that can be harmful or dangerous to others, or the patient encourages or involve others into these situations. This indicator is also denominated by heteroaggressiveness.

- **Self Neglect**: The patient disregards actions of self-care essential to health and maintenance of his/her well-being, such as: hygiene, adequate clothes (for the patient’s age or for the season).

- **Vulnerability**: External factors that a patient can be exposed to and that could be personally harmful, for example: family or social pressure, poverty, homelessness, bullying, among others.

It’s indispensable to collect information about these risk indicators to determine the state of the patient. For each risk previous episodes, current situation and behaviors regarding that risk should be known. We can conclude that GRiST platform uses as input variables of several aspects concerning the user’s state, which are stored securely in a database. It performs an evaluation on the data and produces estimations of the severity of the risk indicators. This process is illustrated in Figure 3.3.

![Figure 3.3: Input and Output variables of GRiST [5]](image)

### 3.6 Risk Assessment in the current instrument

The current hospital's instrument performs the risk assessment for the risk categories of: suicide, self-aggressiveness, heteroaggressiveness, escape, adverse drug reactions and organic pathology. We will only focus on the risk categories of suicide, self-aggressiveness, heteroaggressiveness and escape, since most of the times the remain risks are not taken in consideration by the health professionals when using the current hospital tool.
For each risk indicator/category, there exists a group of questions concerning the factors that may have influence on the risk. All questions on the model are of the type ‘check box’, they are filled in if the patient presents the symptom. The current questionnaire was translated to English and is presented at figure 3.4, the portuguese version currently used in the hospital is presented at figure A.3.

After answering the questionnaire the health professionals follow rules to generate a prediction of the severity level of the risk. These rules consist in a combination of the factors presented in each question that defines the severity of the risk. For each risk category the severity level may vary according to the scale Very High, High, Medium and Low. The rules used for predicting the risk are described at table 3.1, being the numbers the correspondent questionnaire’s question presented on 3.4.

<table>
<thead>
<tr>
<th>Low</th>
<th>Medium</th>
<th>High</th>
<th>Very High</th>
</tr>
</thead>
<tbody>
<tr>
<td>No criteria</td>
<td>1 criterion</td>
<td>1 criterion associated with another criteria</td>
<td>1.2 + 2 + 3</td>
</tr>
<tr>
<td>No criteria</td>
<td>No criteria</td>
<td>2 criteria</td>
<td>3 criteria</td>
</tr>
<tr>
<td>No criteria, except 8</td>
<td>2 criteria</td>
<td>7 + 6 and/or +1</td>
<td>3 + 5 + 8 (or just 8)</td>
</tr>
<tr>
<td>No criteria</td>
<td>No criteria</td>
<td>2 criteria</td>
<td>3 criteria</td>
</tr>
</tbody>
</table>

Children and adolescents that present severity level of ‘Very High’ are the most concerning ones and they need more severe treatment plans and interventions to avoid possible future dangerous situations [13]. A good prediction is critical to guarantee safety and good care; seeing that making an error in this process can bring serious problems due to the patient’s safety and to others, as well as the possibility of children being subjected to a treatment that is not the most appropriate.

All this process of risk management is made by hand by the doctor and nurses. The model is presented on a paper that is filled in whenever a new evaluation is made and is added to the patient’s file.
**Figure 3.4:** Hospital’s current questionnaire English version

<table>
<thead>
<tr>
<th>Identification of Clinical Risk Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Suicide Risk</strong></td>
</tr>
<tr>
<td>1. Suicide attempt</td>
</tr>
<tr>
<td>1.1. Disruptive (impulsive with severe method)</td>
</tr>
<tr>
<td>1.2. Recurrent</td>
</tr>
<tr>
<td>2. Maintenance of Suicidal Ideation</td>
</tr>
<tr>
<td>3. High lethality of the act (aggressive and / or planned method; intention of death expressed or inferred)</td>
</tr>
<tr>
<td>4. Mood disorder, or high impulsivity, psychoses.</td>
</tr>
<tr>
<td>5. Severe family dysfunction and / or family history of t. of suicide or suicide.</td>
</tr>
<tr>
<td>6. Major affective loss due to death or termination of a relationship of fundamental affective support, still felt strongly by the patient.</td>
</tr>
<tr>
<td>7. Serious socialization problems / social isolation.</td>
</tr>
<tr>
<td>9. History of child abuse; Substance use</td>
</tr>
<tr>
<td>10. Male</td>
</tr>
<tr>
<td>Risk (VH/H/M/L)</td>
</tr>
</tbody>
</table>

| **Self-aggressiveness Risk**           |
| 1. History of maintained and recurrent self-harm behaviors. |
| 2. Behaviors of self-aggression and existence of psychotic pathology, personality or mental weakness. |
| Risk (VH/H/M/L)                        |

| **Hetero aggressiveness Risk**         |
| 1. Tense facial expression, loud and fast tone of voice, stare and threatening or avoidance of eye contact with the observer. |
| 2. Tense / threatening posture (wrists, clenched teeth) |
| 3. Excessive motor agitation / aggressive movements (kicks, punches to objects). |
| 4. Aggressiveness and / or verbal threats: (sarcasm, ridicule, less forgiving references to differences, comments of contempt, mistrust, challenge). |
| 5. Dispute and frequent breach of service rules. |
| 6. Delusional perceptions of paranoid nature. |
| 7. Threat of physical aggression directed at objects or people. |
| 8. Aggressiveness expressed, physical, directed at objects or people. |
| Risk (VH/H/M/L)                        |

| **Escape Risk**                        |
| 1. Low consistency in the child’s and / or family’s adherence to the therapeutic project. |
| 2. Previous history of “escapes”       |
| 3. Previous history of additive behavior. |
| Risk (VH/H/M/L)                        |
4 Solution

Contents

4.1 Data Sample ......................................................... 28
4.2 Data Treatment ..................................................... 37
4.3 Decision Trees ....................................................... 38
4.4 The new Questionnaire ............................................. 43
4.5 Arisco System ....................................................... 48
To improve the hospital’s risk assessment process, we propose the development of system ARisCo, that will be responsible for performing this process automatically. The goal of ARisCo is to perform the risk assessment process more efficiently, reducing the time consumed and increasing the precision of the risk assessment.

The development of this system will be separated in three tasks:

- First, to understand the current situation of the hospital’s instrument used to perform the risk assessment, a study focused on the data sample collected from previous patients was made, as described in section 4.1. Also the data sample had to be clean and treated to be used efficiently, as described at section 4.2.

- After having the data sample suitably arranged, a new classification model for the risk assessment was created. Classifiers in form of decision trees were developed to perform the classification of the different risk categories, described in section 4.3. Also, to develop an efficient model for classifying the risk, it was necessary to review the current questionnaire and modify it, remove unused questions, change or add new ones, as described in section 4.4.

- Having the new classification model for the evaluation of the risk, composed by the decision trees and the new revised questionnaire, a web platform was developed to perform the risk assessment process automatically, described in section 4.5.

### 4.1 Data Sample

To collect the hospital’s current instrument evaluations several meetings on the hospital were scheduled. After going through all the old patient’s files, 346 evaluations were collected of patients hospitalized between the years of 2015 and 2017. From those the 346 evaluations, 25 were excluded because they were considered invalid, an evaluation is considered invalid if not all fields are filled in. All the 25 excluded evaluations had responses to the questionnaire’s questions but did not have the final level of risk. After removing the invalid evaluations, the data sample is composed by 321 valid evaluations.

Each evaluation is composed by the patient’s identifier number, evaluation’s date, questionnaire’s answers and evaluation’s result for the risks: Suicide, Self-aggressiveness, Heteroaggressiveness and Escape.

This is the data sample that will be used in the studies performed in this project.

#### 4.1.1 Data Sample’s Demographic Information

The data sample is composed by 321 evaluations with the following characteristics:
• The data sample is composed by 171 patients. A patient can have more than one evaluation during their hospitalization.

• Each patient is identified with a code that is maintained during the hospitalization. Even if the patient is discharged and for some reason is hospitalized again this code will be the same.

• The maximum number of evaluations per patient is 6. There are four patients that have 6 evaluations. Only one of these patients had been hospitalized one time, during 3 months. The other three patients with 6 evaluations were hospitalized more than one time, between 2 or 3 over the 3 years of collected data.

• Ignoring the instances collected that only have a single evaluation, the shortest time of hospitalization is 1 day.

• The longest time of hospitalization is 102 days, almost 3 and half months.

• The average of the hospitalization’s time per patient is 9 days.

• The average of days between evaluations is approximately 10 days.

• The sample is mainly composed by female patients, having 224 female participants and 97 male patients.

Regarding the level of the risks’ predicted on the data sample, table 4.1 presents the number of instances classified for each risk category.

<table>
<thead>
<tr>
<th></th>
<th>Low (number of instances)</th>
<th>Medium (number of instances)</th>
<th>High (number of instances)</th>
<th>Very High (number of instances)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suicide Risk</td>
<td>52</td>
<td>111</td>
<td>99</td>
<td>59</td>
</tr>
<tr>
<td>Self-Aggressiveness Risk</td>
<td>142</td>
<td>71</td>
<td>79</td>
<td>29</td>
</tr>
<tr>
<td>Heteroaggressiveness Risk</td>
<td>221</td>
<td>55</td>
<td>12</td>
<td>33</td>
</tr>
<tr>
<td>Escape Risk</td>
<td>199</td>
<td>87</td>
<td>21</td>
<td>14</td>
</tr>
</tbody>
</table>

It is possible to conclude from the data sample’s demographic information that the data is not correctly balanced regarding the gender of the patients and the final risks levels. Regarding the patient's gender, there are over the double of female patients than male hospitalized between the years of 2015 and 2017. In all the risk categories the level of risk with less instances is the 'Very High', in escape risk there are only 14 patients with a very high risk among the 321 instances in the data sample. These values can bring problems when using the dataset because the 'Very High' level of the risks is the class that is more worrying to correctly classify but it is also the class that has the less instances, it is the minority class.
4.1.2 Current Instrument Evaluation

On the first analysis of the data collected it was possible to detect that not all the risk’s predictions were made according to the hospital’s instrument as it was expected to be. It is important to verify how many evaluations were made according to the instrument and how many were made based on the health professionals intuition, ignoring the rules.

The rules of the current instrument can be treated as a classifier described in section 2.3, and so it can be evaluated with the performance metrics described in section 2.4.

This classifier was developed and run on the data sample in order to check its accuracy, meaning what is the percentage of times the instrument were used to calculate the risk versus the number of times the medical team used their intuition and knowledge to evaluate the patients, not using the instrument’s rules.

<table>
<thead>
<tr>
<th>Table 4.2: Precision of the current questionnaire rules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision of the Instr. Classifier (Low, Medium, High and Very High)</td>
</tr>
<tr>
<td>Suicide Risk</td>
</tr>
<tr>
<td>Self-aggressiveness Risk</td>
</tr>
<tr>
<td>Heteroaggressiveness Risk</td>
</tr>
<tr>
<td>Escape Risk</td>
</tr>
</tbody>
</table>

In table 4.2 the values of classifier’s precision developed with the rules of the current hospital’s instrument are represented. The accuracy of the original risk assessment performed at the hospital with 4 levels for each risk was calculated, and it is presented in the first column. Since the level of risk that is more concerning is the ‘Very High’ it was calculated the accuracy of the hospital’s rules classifier with only on 2 risk’s levels: ‘not Very High’ (Low, Medium and High) and ‘Very High’, as presented in the second column.

Before performing this study, it was expected that the accuracy would be 100% for every risk category, meaning that the health professionals always used the current instrument when performing a patient’s evaluation. Analyzing the table 4.2 it is possible to see that only 58.4% of the evaluations were made using the instrument’s rules for predicting the risk of suicide. For the self-aggressiveness risk 93.5% of the evaluations were predicted using the rules, 89.1% for the heteroaggressiveness risk and 93.8% for the escape risk.

Since the most important level of risk being classified is the ‘Very High’ the metric of sensitivity for the ‘Very High’ class was calculated as presented in the table 4.3 where only the ‘Very High’ instances in the dataset are considered. Observing tables 4.2 and 4.3 it is possible to conclude that the hospital rules present higher metrics values when classifying with only 2 risk’s levels: ‘not Very High’ and ‘Very High’ than when classifying with the 4 risk’s levels which is the scale currently used. Although the accuracy metric with 2 risk’s levels is high it does not mean that the current instrument performs a correct
Table 4.3: Sensitivity class Very High of the current questionnaire rules

<table>
<thead>
<tr>
<th>Risk</th>
<th>Sensitivity of the Rules Classifier (only ‘Very High’ instances considered)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suicide Risk</td>
<td>42.4%</td>
</tr>
<tr>
<td>Self-aggressiveness Risk</td>
<td>79.3%</td>
</tr>
<tr>
<td>Heteroaggressiveness Risk</td>
<td>84.8%</td>
</tr>
<tr>
<td>Escape Risk</td>
<td>71.5%</td>
</tr>
</tbody>
</table>

risk’s classification, it still presents poor values classifying the ‘Very High’ class, for the suicide rules only 42.4% of the Very High instances are well classified.

We also calculated the confusion matrices for each risk category regarding the current evaluation instrument.

In the tables 4.4, 4.5, 4.6, 4.7 the confusion matrices for each one of the risks being analyzed are presented.

Table 4.4: Questionnaire’s Rules Confusion Matrix for Suicide Risk

<table>
<thead>
<tr>
<th>Hospital Rules’ Evaluation</th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
<th>Very High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample’s Evaluation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>36</td>
<td>14</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Medium</td>
<td>1</td>
<td>28</td>
<td>82</td>
<td>0</td>
</tr>
<tr>
<td>High</td>
<td>0</td>
<td>0</td>
<td>98</td>
<td>1</td>
</tr>
<tr>
<td>Very High</td>
<td>0</td>
<td>0</td>
<td>34</td>
<td>25</td>
</tr>
</tbody>
</table>

Analyzing the table 4.4 regarding the suicide risk we can observe that 187 of the instances are correctly classified highlighted in green, 36 instances that were classified as Low by the health professionals are also classified as Low using the instrument’s rules, 28 classified as Medium, 98 as High and 25 as Very High. These are the evaluations that the classification performed by the hospital instrument’s rules match the evaluation made by the doctors and nurses, presented in the sample collected. The values outside the matrix’s diagonal are the incorrectly classifications by the rules. It’s possible to see one particular worrying number, 34 evaluations that the health professionals considerer being of the risk ‘Very High’ are evaluated by the questionnaire’s rules by being only a ‘High’ risk. When the instrument’s rules predict a risk level lower than the professionals think it should be is worrying because the child will not receive the attention according with it’s real risk’s level.

Table 4.5: Questionnaire’s Rules Confusion Matrix for Self-aggressiveness Risk

<table>
<thead>
<tr>
<th>Hospital Rules’ Evaluation</th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
<th>Very High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample’s Evaluation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>140</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Medium</td>
<td>1</td>
<td>64</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>High</td>
<td>0</td>
<td>2</td>
<td>73</td>
<td>4</td>
</tr>
<tr>
<td>Very High</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>23</td>
</tr>
</tbody>
</table>
Regarding the self-aggressiveness risk evaluations presented in table 4.5, it is possible to observe that there are some evaluations given by the instrument’s rules that are concerning. Although there exist 300 evaluations where the health team intuition and the hospital instrument’s rules match (highlighted in green), there are 6 worrying evaluations. It is possible to see that 1 instance was evaluated by the doctors and nurses as ‘Very High’ but the questionnaire’s rules evaluated it with a ‘Low’ risk, 2 instances were evaluated by the rules as being ‘Medium’ when the health team considered them ‘Very’ High and 3 instances were evaluated by the rules as being ‘High’ but the health professionals considered them as being ‘Very High’.

**Table 4.6: Questionnaire’s Rules Confusion Matrix for Heteroaggressiveness Risk**

<table>
<thead>
<tr>
<th>Sample’s Evaluation</th>
<th>Hospital Rules’ Evaluation</th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
<th>Very High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td></td>
<td>210</td>
<td>4</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Medium</td>
<td></td>
<td>0</td>
<td>38</td>
<td>13</td>
<td>4</td>
</tr>
<tr>
<td>High</td>
<td></td>
<td>1</td>
<td>0</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>Very High</td>
<td></td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>28</td>
</tr>
</tbody>
</table>

**Table 4.7: Questionnaire’s Rules Confusion Matrix for Escape Risk**

<table>
<thead>
<tr>
<th>Sample’s Evaluation</th>
<th>Hospital Rules’ Evaluation</th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
<th>Very High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td></td>
<td>192</td>
<td>6</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Medium</td>
<td></td>
<td>1</td>
<td>83</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>High</td>
<td></td>
<td>2</td>
<td>2</td>
<td>16</td>
<td>1</td>
</tr>
<tr>
<td>Very High</td>
<td></td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>10</td>
</tr>
</tbody>
</table>

We were able to observe that several predictions of the severity level for the risk categories do not match the ones predicted by the current instrument’s rules. For example, several severity levels were classified by the health professionals as ‘Very High’ but following the model’s rules they should be classified as ‘Medium’. This situation occurred because the health professionals took more into account their intuition on the state of the patient than the instrument’s rules. Detecting this situation on the evaluations collected, we concluded that the current model does not follow the health professionals’ intuition concerning the state of the patient. This results in a manipulation of the current model factors in order to obtain the level of severity that the health professionals think is the most appropriate for the patient. There are cases where the health professionals do not even fill in the questionnaire’s questions and only write the severity level of each risk that they think is the most appropriate for the patients, reflecting the invalid instances referenced in the previous section 4.1.1.

Making an average using the values presented on table 4.2, only 83.7% of the 321 instances the risk’s level given by the current hospital rules match the ones given by the health professionals.
4.1.3 Risks’ Correlation

To have a better understanding of the data and the relations within the four risks’ categories some studies were made to identify some interesting relations. Relations between the 4 risks’ categories were analyzed, calculating for each risk ‘Very High’ how many evaluations regarding the other risk categories are also ‘Very High’.

Table 4.8: Relation regarding Very High evaluations between the risks: Suicide, Self-aggressiveness, Heteroaggressiveness and Escape.

<table>
<thead>
<tr>
<th></th>
<th>Suicide Risk VH</th>
<th>Self-aggressiveness Risk VH</th>
<th>Heteroaggressiveness Risk VH</th>
<th>Escape Risk VH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suicide Risk VH</td>
<td>59</td>
<td>13</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Self-aggressiveness Risk VH</td>
<td>13</td>
<td>29</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Heteroaggressiveness Risk VH</td>
<td>2</td>
<td>5</td>
<td>33</td>
<td>8</td>
</tr>
<tr>
<td>Escape Risk VH</td>
<td>3</td>
<td>0</td>
<td>8</td>
<td>14</td>
</tr>
</tbody>
</table>

The conclusion taken from this study, with the results presented at table 4.8, is that the two risks that are more correlated with each other are suicide and self-aggressiveness. Observing table 4.8 it is possible to see that within 59 patients with a ‘Very High’ risk of suicide 13 also present a very high risk of self-aggressiveness, and among 29 patients with a ‘Very High’ risk of Self-aggressiveness 13 also present a ‘very High’ risk of suicide.

This event can happen because of the similarities of the patient’s disease, it is possible to present symptoms that fall both in the ‘Very High’ level of suicide and self-aggressiveness risk.

The questionnaire’s question 3 from the Self-aggressiveness section - "Self-harm behaviors with risk of suicide", is used to calculate the level of self-aggressiveness risk and it takes in account suicide symptoms, which show that the symptoms regarding these two risks are correlated.

4.1.4 Questionnaire Questions’ Relations

Besides the study made on the correlations between the final risk’s levels, an analysis on the 25 questionnaire’s questions was also made. The goal here was to discover possible relations among the questions.

To discover these relations the equation of Pearson Correlation Coefficient [21] described in section 2.8.1 was used. This equation receives two arrays of data, meaning, two columns with the responses to two questionnaires’ questions, returns a number between -1 and 1, called the correlation coefficient of those questions.

When the correlation coefficient is 1 it indicates that there exists a positive strong relation between the questions, -1 indicates there exists a negative strong relation. When the value is 0 indicates that there are no relations between the questions. From this study we concluded:
• Question 7-“Threat of physical aggression directed at objects or people.” and Question 8-“Aggressiveness expressed, physical, directed at objects or people.” from the section of heteroaggressiveness, have a correlation coefficient of 0.731.

• Question 1.1-“Suicide attempt: disruptive(impulsive with severe method)” and Question 3-“High lethality of the act (aggressive and/or planned method; Intention of death expressed or inferred).” of suicide risk have a correlation coefficient of 0.6455.

We can conclude that these pairs of questions have a strong positive correlation, following the rule of Tumb presented in table 2.3, which means that when a answer to one is “Yes” the answer to the other is also “Yes”.

There are many questions that do not have any kind of relation following the rule of Tumb, Question 2-“Previous history of "escapes"” from escape section and Question 1-“History of maintained and recurrent self-harm behaviors.” from self-aggressiveness section have a correlation coefficient of 0.0033.

4.1.5 Questionnaire Questions and Risk’s Relations

To understand the potential relationships between the questionnaire’s questions and the final level of risk, several studies were made. To identify these relations two techniques were used: Pearson correlation coefficient presented on section 4.1.5.A and LR models presented on section 4.1.5.B.

4.1.5.A Correlation coefficient

A study between the questionnaire’s questions and the final four risk’s categories was made to find the relationships between the current hospital’s instrument entries (questions) and the its exits (risk’s levels). Similarly to the study of section 4.1.4, the equation of the Pearson Correlation Coefficient [21] described in section 2.8.1 was used.

The results of this analysis are presented in table B.1. Observing at the table B.1 is possible to conclude the strongest relationships are:

• Heteroaggressiveness risk category and Question 8-“Aggressiveness expressed, physical, directed at objects or people.”, present a correlation coefficient of 0.8167, following the rule of Tumb described on table 2.3, this value shows a strong positive correlation.

• Suicide risk category and questions 2-“Maintenance of Suicidal Ideation” and 3-“High lethality of the act (aggressive and/or planned method; intention of death expressed or inferred).”, present the correlation coefficient of 0.6048 and 0.6077 respectively.

• Escape risk category has a correlation coefficient of 0.5191 with Question 3-“Previous history of additive behavior”
All these correlation coefficients are positive, showing that exists a positive relation between the pairs (question/risk category) being analyzed. This means that when the answer to this questions is “Yes” the final risk is ‘Very High’, most of the times. We can also conclude that the higher correlations happen in the questions of the same risk’s category, the correlation coefficients between questions of a different risk’s category are close to 0, meaning that there is no relation of an individual question with risk’s of a different category.

4.1.5.B Logistic regression

Similar to the previous section 4.1.5.A a study between the questionnaire’s questions and the final four risk’s categories was made to find existing relationships, but using the binary LR technique described at section 2.8.2. A LR model was developed to each risk category with the levels ‘not Very High’ and ‘Very High’, using the questions regarding risk’s categories. LR uses the data sample to make estimations creating a regression model which will be responsible for determining which entries (questions) have a relation with the exit (risk category). With the resulting LR models it was possible do conclude:

• Regarding the suicide risk, the questions that present the strongest association are: Question 2- “Maintenance of Suicidal Ideation” and Question 3-“High lethality of the act (aggressive and / or planned method; intention of death expressed or inferred)”.

• Regarding the self-aggressiveness risk, the questions that have the strongest association are: Question 2-“Behaviors of self-aggression and existence of psychotic pathology, personality or mental weakness.” and Question 3-“Self-harm behaviors with risk of suicide”.

• The question with the strongest relation with heteroaggressiveness risk is Question 8-“Aggressiveness expressed, physical, directed at objects or people”.

• The questions that have the strongest relation with the final risk level regarding the escape risk are Question 2 -“Previous history of ‘escapes’” and Question 3-“Previous history of additive behavior”.

Although the hospital questionnaire’s rules are not used every time to calculate the level of the risk categories, the questions mentioned are the ones that have impact on the rules used by the current instrument to determine the ‘Very High’ risk. For example, the rule to assign a ‘Very High’ risk on the risk category of heteroaggressiveness is having the criteria of Question 8-“Aggressiveness expressed, physical, directed at objects or people.”, so it was expected that this question would have a strong relation with the final level of risk. Questions 2 and 3 from suicide also are the ones used by the rules to assign a ‘Very High’ risk so it was also expected that they would have a strong relation with the final level of suicide risk.

1Study performed in collaboration with Professor Dina Salvador, School of Technology, Polytechnic Institute of Setúbal
Besides finding the most significant variables, the LR model also shows some questions that were not relevant:

- In the suicide risk section questions 8-“Feelings of despair, lack of hope and low self-esteem” and 9-“History of child abuse; Substance use” were categorized by the model as not being statistically significant regarding the final risk of suicide.

- Regarding the heteroaggressiveness risk, questions 2-“Tense/threatening posture (wrists, clenched teeth)” and 7-“Threat of physical aggression directed at objects or people” were considered by the regression model as not being statistically significant for the final level of risk.

These conclusions were expected since the questions mentioned before are not used by the current instrument rules to calculate any level of risk, it is natural that they wouldn’t have influence in the final level of risks.

4.1.5.C Main conclusions

Observing the values obtained from both studies, one using the Pearson Correlation Coefficient the other using a LR model, some of the final results are similar, allowing us to take main conclusions, regarding each risk category:

- Regarding the suicide risk, both approaches show that the questions with a strongest association with this risk category are: Question 2-“Maintenance of Suicidal Ideation” and question 3-“High lethality of the act (aggressive and/or planned method; intention of death expressed or inferred).”.

- Regarding the self-aggressiveness risk, from the LR approach we concluded that the questions with the strongest association are: Question 2-“Behaviors of self-aggression and existence of psychotic pathology, personality or mental weakness.” and Question 3-“Self-harm behaviors with risk of suicide”.

- In both approaches, Question 8-“Aggressiveness expressed, physical, directed at objects or people.” is the one with the strongest relationship with the risk category of heteroaggressiveness.

- Also in both studies Question 3-“Previous history of additive behavior.” revealed a strong relation with the risk category of escape.
4.2 Data Treatment

4.2.1 Final Risks’ Levels

The data sample collected described in section 4.1 contains the risks categories of Suicide, Self-aggressiveness, Heteroaggressiveness and Escape. Each risks level can vary between: Low, Medium, High and Very High. According to the health professionals participating in this project, a child is hospitalized only if it presents a Very High risk in any risk category, if it does not is not hospitalized. With this fact it was concluded that there is not necessary having 4 different risk’s levels, it is only relevant to know if a child has a Very High level at any risk’s category so it is/remains hospitalized or not. In order to simplify the new instrument being developed it was decided by the team to only have 2 levels of risk instead of 4: not Very High risk (encompassing Low, Medium and High) and Very High risk. With this decision the data sample was modified, the current number of risks’ classifications within the 321 instances is presented at table 4.9:

<table>
<thead>
<tr>
<th></th>
<th>not Very High (Low, Medium, High)</th>
<th>Very High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suicide Risk</td>
<td>262</td>
<td>59</td>
</tr>
<tr>
<td>Self-Aggressiveness Risk</td>
<td>292</td>
<td>29</td>
</tr>
<tr>
<td>Heteroaggressiveness Risk</td>
<td>288</td>
<td>33</td>
</tr>
<tr>
<td>Escape Risk</td>
<td>307</td>
<td>14</td>
</tr>
</tbody>
</table>

Table 4.9: Data sample’s risk classification converted to not Very High and Very High

4.2.2 Balancing Data

As already concluded in section 4.1.1, the data sample collected is not balanced. Even with the modifications of the risks’ levels described in the previous section 4.2.1, the amount of instances classified as ‘not Very High’ are a lot higher than the ones classified as ‘Very High’. For a classifier to perform well, the data must be balanced, meaning there should exist as many instances from one class as there are from the other. Creating classifiers for an unbalanced data will have a negative impact on their performance, because the classifier can ignore the instances from the minority class since evaluating the ones from the majority class will already have a positive performance. This is especially worrying because the minority class in the data sample is the ‘Very High’ risk, where the correct classification is more important, because classifying a child with a ‘not Very High’ risk when should be a ‘Very High’ risk is a dangerous mistake. To resolve this issue, the SMOTE [20] technique described in section 2.7.1 was used. This algorithm will create new instances on the minority class (Very High) in order to match the number of instances in the majority class (not Very High).
4.3 Decision Trees

Classifiers in form of decision trees were chosen to perform the risk assessment for ARisCo system, because it is easy for the health professionals to understand the classification process when it is represented through a tree. In section 4.3.1 is described all the steps taken for developing the decision tree classifiers. The final decision tree’s classifiers are presented at section 4.3.3.

The decision tree’s development process was made in parallel to the new questionnaire’s creation described in section 4.4 since alterations in one of the processes produce alterations in the other.

4.3.1 Decision Trees Classifiers’ Development

A decision tree was developed for each of the risks categories: Suicide, Self-aggressiveness, Heteroaggressiveness and Escape. All the decision trees were created using the C4.5 algorithm [17] described in section 2.6. We used a balanced version of the data sample collected, as described in section 4.2.2 and cross-validation as described in section 2.3.

The decision trees are composed by the questionnaire’s questions and they have as class the two different possible values for the risks categories: ‘Very High’ and ‘not Very High’.

The data sample is not correctly balanced, having more instances classified as ‘not Very High’ than ‘Very High’. The data sample was balanced separately for each one of the risk’s categories: Suicide, Self-Aggressiveness, Heteroaggressiveness and Escape, since the ratio between the ‘Very High’ and ‘not Very High’ classes vary between the categories. After applying the SMOTE algorithm, there exist 4 different datasets, each one having the number of instances balanced according to its risk category. Each dataset will be used for the development of the correspondent decision tree risk’s category.

4.3.2 Risk category’s counters

Besides the questionnaire’s questions, counters were also added to the data sample to improve the performance of the decision trees and also to decrease their size. These counters were already used in the current hospital’s instrument in the categories of self-aggressiveness and heteroaggressiveness since the rules for classifying these risk were: no criteria - Low, 1 criteria - Medium, 2 criteria - High and 3 criteria - Very High.

The following counters were added to the data sample:

- **Suicide Counter “number” or more Yes’ responses**: Takes the values: Yes or No. Yes if the number of Yes’ answers on the Suicide section’s questions is equal or bigger than the counter’s number, No if the opposite. There are 6 counters for the Suicide section: 0 or more, 1 or more, 2 or more, 3 or more, 4 or more and 5 or more.
• **Self-aggressiveness Counter "number" or more Yes' responses:** Takes the values: Yes or No. Yes if the number of Yes’ answers on the Self-aggressiveness section’s questions is equal or bigger than the number of the counter, No if the opposite. There are 2 counters for the Self-aggressiveness section: 0 or more and 1 or more.

• **Heteroaggressiveness Counter "number" or more Yes' responses:** Takes the values: Yes or No. Yes if the number of Yes’ answers on the Heteroaggressiveness section’s questions is equal or bigger than the number of the counter, No if the opposite. There are 4 counters for the Heteroaggressiveness section: 0 or more, 1 or more, 2 or more and 3 or more.

• **Escape Counter "number" or more Yes' responses:** Takes the values: Yes or No. Yes if the number of Yes’ answers on the Escape section’s questions is equal or bigger than the number of the counter, No if the opposite. There are 2 counters for the Escape section: 0 or more and 1 or more.

The development of the decision trees went through several steps: first the trees were created with the raw data sample collected. Then the balanced data sample was used and afterwards pruning techniques in order to improve the performance and size of the trees.

Several tests were made until the best performance metrics versus size of the tree were achieved. In these tests different input values were given to C4.5 algorithm to perform tree’s pruning.

### 4.3.2.A Pruning techniques performed

Pruning was used on decision trees with the goal of reducing their size without decreasing their performance. To perform pruning, two approaches were used: increase of the number minimal instances per each tree leaves and increase/decrease of the confidence factor given as input parameters to the C4.5 algorithm.

• **Minimal instances per leaf:** By increasing the minimum number of instances per tree’s leaf we are establishing that the decision tree cannot have leaves that categorize few instances. When setting the number of minimum instances as 10, there will be no tree leaf with less than 10 instances. When increasing this value the algorithm will create a smaller and more compact tree and according to the data sample being used the accuracy and sensitivity metric can also increase.

• **Confidence factor:** The confidence factor will establish the amount of pruning performed on the tree. With lower values, there will be a heavy pruning performed on the tree. If the value of this parameter decreases the pruning performed in the tree will decrease. If this value is 0, there will be no pruning performed. The pruning using this parameter is made decreasing the classification errors in each tree node.
Several trees were created for each risk category changing these two parameters until the best performance metrics and tree's size were achieved.

4.3.2.B Input questions

Besides the pruning techniques performed on the trees also the questions given to input were changed along the several tests made. The questionnaire have several questions that have a strong impact on the level of risk result and others that do not, as shown on the studies made described in section 4.1. The manipulation made by the doctors and nurses on the current hospital's instrument made some questionnaire's questions have strange and not correct classification on the tree's leaves. It was necessary to remove questions as input to the c4.5 algorithm to have trees that made sense for the health professionals. For example, in the study described in section 4.1.5.A, one of the questions that have the strongest relationship with the final level of risk of self-aggressiveness is question 3 - "Self-harm behaviors with risk of suicide", but it was decided by the health professionals that this question was not necessary to perform the evaluation, since behaviors related to suicide were already being evaluated on the questions related to suicide.

4.3.2.C Cross Validation

K-fold cross validation technique described at section 2.3 was used to develop the decision tree's classifiers, since the data sample was not very large to be divided in a training and test set. The number of folds was changed several times until the best performance for each tree was found. In all the final decision trees, cross-validation was used with the value of 15 folds, because it was the value that allowed to achieve better performance results.

4.3.2.D Health professionals evaluation

The decision trees classifiers were analyzed by the health professionals involved in this project to guarantee that they were classifying the risks according to their knowledge and intuition. It was possible to conclude that due the manipulations performed on the questionnaire's questions, some not important questions for the risk assessment were used incorrectly in the decision trees. With the help from the nurses involved in the project and their approval these questions were removed from the trees and debating with all the participants involved, it was possible to achieve 4 classifiers that perform well and were accepted by the health professionals to perform the risk assessment. Although these tree classifiers were accepted by the health professionals there is no guarantee that they will perform well in the real time evaluations with different nurses and doctors using them.
4.3.3 Final Decision Tree’s classifiers

After all the tests performed to create the decision trees for classifying the risk, the final 4 decision trees’ classifiers that will be used in ARisCo system are presented in the following sections.

4.3.3.A Suicide Risk Tree

The final tree regarding the suicide risk was created using the data sample balanced for these category, composed by 524 evaluations: 262 instances classified as 'Very High' and 262 classified as 'not Very High'.

Pruning was used by establishing that the minimum instances by tree’s leaf must be 10 and having a confidence factor of 0.3. The final tree regarding the risk of suicide is presented in the image 4.1. The performance tree’s metrics are presented in table 5.1 at section 5.2.

Figure 4.1: Final decision tree for the Risk of Suicide

4.3.3.B Self-Aggressiveness Risk Tree

The final tree regarding the Self-Aggressiveness risk was created using the data sample created for the this risk category, composed by 584 evaluations: 292 instances classified as Very High and 292 classified as not Very High.

Pruning was used by establishing that the minimum instances by tree’s leaf must be 28 and having a
confidence factor of 0.3. The final tree regarding the risk of Self-Aggressiveness is presented in the image 4.2. The tree’s performance metrics are presented in table 5.1 at section 5.2.

**Figure 4.2: Final decision tree for the Risk of Self-Aggressiveness**

![Decision Tree](image)

### 4.3.3.C Heteroaggressiveness Risk Tree

The final tree regarding the Heteroaggressiveness risk was created using the data sample balanced for this category, composed by 576 evaluations: 288 instances classified as 'Very High' and 288 classified as 'not Very High'. Pruning was used by establishing that the minimum instances by tree’s leaf must be 12 and having a confidence factor of 0.3. The final tree regarding the risk of heteroaggressiveness is presented in the image 4.3. The performance metrics of the tree are presented in table 5.1 at section 5.2.

**Figure 4.3: Final decision tree for the Risk of Heteroaggressiveness**

![Decision Tree](image)
4.3.3.D Escape Risk Tree

The final tree regarding the escape risk was created using the data sample balanced for the escape risk category, composed by 614 evaluations: 307 instances classified as 'Very High' and 307 classified as 'not Very High'.

Pruning was used by establishing that the minimum instances by tree's leaf must be 9 and having a confidence factor of 0.3. The final tree regarding the risk of escape is presented in the image 4.4. The performance metrics of the tree are presented in table 5.1 at section 5.2.

4.4 The new Questionnaire

The hospital's current questionnaire was subjected to several changes to create a smaller and improved version to be used in the ARisCo system. As mentioned before, the process of creating a new questionnaire was made in parallel to the development of the new decision tree classifiers responsible for performing the risk assessment, described at section 4.3, because the changes in the questionnaire will affect the decision trees development.

The first big change in the questionnaire was removing the risks sections. In the current questionnaire, each risk had its own questions and those questions were used only to calculate the risk for one specific category. The new questionnaire does not have sections, all questions can be used to calculate any risk's category. This change made us achieve better performance results in the decision trees and reduce the size of the questionnaire. The tree classifiers can use questions from other risks categories to perform the risk assessment.

With the help of the health professionals involved in this project it was possible to remove questions that
were not needed to perform an accurate risk assessment, questions that were duplicated were jointed, questions were reformulated to a more up-to-date writing and also new questions were added. The following sections describe all the changes made on the initial questionnaire.

### 4.4.1 Questions’ Removal

The first target questions to be removed from the questionnaire were the ones that do not appear in any of the decision trees classifiers, because if the decision trees can classify the risk not using that questions, they are not considered relevant. The correlation studies made in section 4.1.5 will not be blindly used for the removal of questions because they only show the correlation of a single question to the final risk. A single question can not have a strong impact on the risk but a group of several questions can. The decision tree classifiers show these strong correlations, since they do not classify the patient using a single question but using a set of questions.

A summary of questions being used at the final decision tree classifiers is presented in table 4.10. This table was discussed with the health professionals in order to achieve a consensus of which questions should in fact be removed or not.

On table 4.10 it is possible to observe that 6 questions are not used in any of the decision trees, thus allowing us to reach the conclusion that they are not relevant to perform the risk assessment. In discussion with the health professionals it was agreed that in fact these questions were not so important to calculate the risk.

The questions removed are:

- **Question 3 from self-aggressiveness section** -“Self-harm behaviors with risk of suicide”, was removed because the health professionals considered that the subject evaluated in this question was already contemplated in the questions of the suicide section.

- **The questions from the heteroaggressiveness section**, 3- “Excessive motor agitation / aggressive movements (kicks, punches to objects).”, 5-“Dispute and frequent breach of service rules” and 6- “Delusional perceptions of paranoid nature”, were removed because they are not used in the trees and the health professionals agreed that the remain questions were enough to perform a accurate risk assessment.

- **Question 2 from heteroaggressiveness section** “Tense / threatening posture (wrists, clenched teeth)”, despite of not being used in the classifiers the health professionals considered it important. Since the subject of the question is the same as question 1 of the same section - “Tense facial expression, loud and fast tone of voice, stare and threatening or avoidance of eye contact with the observer.”, it was decided to join these two questions.
Table 4.10: Questions being used on the final decision trees

<table>
<thead>
<tr>
<th></th>
<th>Suicide Risk Tree</th>
<th>Self-Aggressiveness Tree</th>
<th>Heteroaggressiveness Tree</th>
<th>Escape Tree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sui. 1.1</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sui. 1.2</td>
<td>Counter</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sui. 2</td>
<td>Counter</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sui. 3</td>
<td>Counter</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sui. 4</td>
<td>Counter</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sui. 5</td>
<td>Counter</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sui. 6</td>
<td>Counter</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sui. 7</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sui. 8</td>
<td>Counter</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sui. 9</td>
<td>Counter</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sui. 10</td>
<td>Counter</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self. 1</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self. 2</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self. 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hetero. 1</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Hetero. 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hetero. 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hetero. 4</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Hetero. 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hetero. 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hetero. 7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hetero. 8</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Esc. 1</td>
<td>Counter</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Esc. 2</td>
<td>Counter</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Esc. 3</td>
<td>Counter</td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

- Question 7 from section of heteroaggressiveness - "Threat of physical aggression directed at objects or people." was removed because it was already contemplated in question 8 from the same section - "Aggressiveness expressed, physical, directed at objects or people". This decision was also supported by the correlation study made on the questionnaire’s questions described in section 4.1.4, These two question were the ones with the highest correlation factor meaning their answers are very similar. With these facts this two questions were jointed into one.

- Regarding the questions, 4 from suicide risk - "Mood disorder, or high impulsivity, psychoses." and 2 from escape section - "Previous history of "escapes”", although being used, not in a tree node but in the counter of suicide, they were considered not relevant for the risk assessment and complicated to answer to the health professionals, so they were also removed from the questionnaire.

4.4.2 Questions Divided

Some questions that approach more than one subject were divided, the health professionals considered that it is more correct to treat the subjects separately:
• Question 5 from suicide section- "Severe family dysfunction and/or family history of t. of suicide or suicide" was divided into: “Severe family dysfunction” and “Family history of t. of suicide or suicide”.

• Question 9 from suicide section- "History of child abuse; Substance use" was divided into: “History of child abuse” and “Substance use”.

4.4.3 Duplicate Questions

Some questionnaire’s questions were considered duplicated. Although they are not written the same way, they deal with the same subject. These questions were jointed:

• Question 3 from suicide risk-“High lethality of the act (aggressive and / or planned method; intention of death expressed or inferred).” was considered by the health professionals a duplication of the question 1.1 from suicide risk-“Suicide attempt: Disruptive (impulsive with severe method)”. This is also supported by the correlation study described in section 4.1.4, where these two questions presented a significant correlation coefficient. Question 3 was removed and Question 1.1 remained in the questionnaire.

• Question 3 from escape section-“Previous history of additive behavior.” was also considered a duplication of question 9 that was decomposed as mentioned in section 4.4.2 -“Substance use”.

4.4.4 Questions reformulated

The text of some questions was also reformulated, so that its meaning would be less confusing and also so as to update some health-related vocabulary. The following questions were updated:

• Question 1.1 from suicide-“Suicide attempt:Disruptive (impulsive with severe method) ” was updated into “Suicide attempt:Severe (egg hanging, drowning, firearm)”

• Question 5 from suicide section after divided: “Severe family dysfunction ” and “family history of t. of suicide or suicide”. Its sub-questions were changed to: “History of child abuse” and “Unprotected child (egg. Absence of reference figure, family / institutional)”.

• Question 6 from section suicide-“Major affective loss due to death or termination of a relationship of fundamental affective support, still felt strongly by the patient” was changed into: “Significant recent loss (egg mourning, material loss)”.

• Question 9 from suicide section after being divided into: “History of child abuse” and “Substance use” was updated to: “History of child abuse” and “Abuse of alcohol or substances”.

46
4.4.5 New questions and New risk

After making all the necessary changes on the questionnaire the health professionals felt the need to add new questions they considered important to perform a correct risk assessment. These questions will not be included in the decision trees classifiers because there exist no data of them, and is not possible to know how they will be used in the classifiers. The answers to these questions will be collected and in the future a new analysis will be made to discover where they fit best.

The new questions added to the questionnaire are:

• "Structured suicide plan (ex. Farewell letter, preparations)".

• "Child does not verbally communicate".

• "Risks associated with organic pathology (epilepsy, eating disorders, encephalitis)." ¹

Besides the new questions added the health professionals considered important to add a new risk category to be evaluated: "Risk of Clinical Pathology." This new risk is calculated with the last question added ¹. When its answer is Yes the value of this risk is Very High, if the opposite occurs the value of the risk is not Very High.

4.4.6 Final Questionnaire

After all the changes, a new questionnaire was created. It is a more correct, intuitive and small one, allowing the doctors and nurses to perform risk assessment more efficiently. The final questionnaire was translated into English and is presented in figure 4.5. The portuguese version of the final questionnaire, which will be used in the hospital is presented on figure A.1.
4.5 Arisco System

Having new model for risk assessment composed by the decision trees and the new questionnaire, that will offer a more efficient way to perform the children evaluation and also validate the decision trees classifiers, a web system was developed to be used at the hospital, the ARisCo system. The goal is that the ARisCo system will, in the future, replace the current pen & paper instrument current being used. ARisCo will save the patient's information on a data base and run the decision trees retrieving to the health professionals the risks for that patient. Besides replacing the current instrument this system will also be used to evaluate of the risk assessment performed by the tree classifiers. After the new questionnaire's questions the system will have a section where the professionals will perform their own risk assessment based on their knowledge and intuition. The data will be stored and in the future it will be possible to compare the risks levels given by the tree classifiers with the ones given by the health professionals.

4.5.1 System’s Functionalities and Architecture

The web system was developed to be run via browser to be accessed on the hospital’s computers and maybe in the future on hospital’s tablets.
The system is composed by several web pages that were developed using HTML, CSS and Javascript, available to be accessed by a link. The server side of the system developed using Nodejs, is located in a virtual server: EvenNode, always available to receive requests from the client side. For storing the data, is used a mysql database hosted in Instituto Superior Técnico (IST).

When a health professional accesses the ARisCo system, the init page is displayed, illustrated at figure 4.6. The system offers three functionalities, described in the next sections.

Figure 4.6: System Init Page

4.5.1. A Risk Evaluation functionality

This is ARisCo’s main functionality and it allows the user to perform a risk evaluation for a patient. After the health professional identifies the patient with its identification number, the questionnaire illustrated at figure 4.5 is displayed. The page displayed to the health professionals is presented at figure A.4 at the attachments.

The questionnaire’s questions are stored in the database, if there is a need to change any question or add new ones it is possible to change them using the database. It is also possible to choose which questions are going to be displayed on the page and change their order.

As shown in figure 4.7, when the health professional reaches the system’s init page, a request is sent to the server asking for the questions to display. The server connects to the database, gets the questions and sends them back to the client to be displayed, action orange on the figure 4.7. This request is made using the identification number of the patient because in a first evaluation all questionnaire’s questions are displayed, but on posterior evaluations, only the questions whose answer can change during the child hospitalization are displayed. For example, the answer to Question 9- “History of child abuse”, will always be the same, it will not change during the patient’s hospitalization. This kind of questions are only displayed to the health professionals one time.

Once the questions are displayed the health professionals must answer to all and also perform a risk assessment according with their knowledge and intuition, for the risks of Suicide, Self-Aggressiveness, Heteroaggressiveness, Escape and Risk of Clinical Pathology.

The health professionals are asked to perform their own risk assessment with the goal of, after collecting
some evaluations, compare the risk’s levels predicted by the decision trees classifiers to the ones given by the health professionals.

After the data is filled in, the classifier algorithm is run and all the information is sent to the server to be stored in the database, action marked blue on the figure 4.7. After all the data is saved a resume page of the risk evaluation is presented, as illustrated at figure 4.8.

The decision trees classifiers risk’s evaluation is not displayed to the user to avoid influencing the health professionals questionnaire’s responses and avoid manipulations.

Figure 4.8: ARisCo system resume page
4.5.1.B  Change fixed question response functionality

As described in the previous functionality details on section 4.5.1.A, there exists fixed questions. These questions are only displayed in the first patient's evaluation, since their answers do not change during the patient's hospitalization, they are factors that are a part of the patient's life. Since these questions are no longer available on the questionnaire, this functionality was created in case of a mistake answering them in the first time.

With the patient's identification number, identifying the number of the fixed question and choosing the new answer, the response to the question will be modified in all entries of that patient stored in the database, as described in figure 4.9.

Figure 4.9: System's architecture of functionality Change fixed question

4.5.1.C  Check last evaluations functionality

This last functionality was purposed by the health professionals. It was necessary to check when the last evaluations were made to a specific patient. The evaluations are not all performed by the same health professional, this functionality aims to avoid duplicated evaluations and also to avoid forgetting to perform one. Using patient's identification number the last three evaluation's are displayed. The system's architecture to perform this functionality is illustrated at figure 4.10.

With this functionality, when a health professional starts performing a risk evaluation, he/she will check if none of their colleges had already performed the risk assessment and also check if some evaluation is missing.

4.5.2  Decision Tree Algorithm

When the questionnaire's answers are submitted to the server the algorithm responsible to run the decision trees classifiers for the 4 risks categories is called to perform the risk assessment. The 4 decision trees are stored in JSON format on the client side. An example of the structure of the decision tree's JSON file is presented on C.1 on attachments, containing the decision tree for suicide risk. In the future, to perform an alteration on the trees is only needed to change the decision tree's
structure on the JSON files. The decision tree’s algorithm is responsible for consulting the JSON files and using the questionnaire’s answers, predict the levels of risk. The pseudo code of this algorithm is presented at section C.1 on attachments.

4.5.3 Database’s Structure

The database is structured with 3 different tables to efficiently store all the necessary data. The architecture of the database is presented in the figure 4.11.

All the information regarding the questionnaire’s questions to be displayed on the ARisCo web site is stored in the table questions, this table has the following columns:

- **Id**: Table’s primary key, integer identifying the question.
- **questOrder**: Integer defining the position of the question on the questionnaire.
- **Description**: String with the question’s text to be displayed.
- **Active**: Boolean that defines if the question is active or not, if it is displayed or not.
• **Category**: String representing the question’s category: Suicide, Self-aggressiveness, Heteroaggressiveness, Escape, Pathology and New.

• **separatedFrom**: Integer representing the id of the question that were separated from. This value is used for the questions that were divided (necessary to calculate the response for the decision tree’s paths).

• **subQuestion**: Boolean identifying if it is a single question or is a subquestion of another.

• **sectionId**: Integer can take the value 1 or 2, 1 if the question belongs to the fixed question’s category, it is only showed to the client in the first evaluation. 2 if the question belongs to the section of questions that is always displayed to the client.

The table Responses stores all the questions answered for the evaluations, it has the following columns:

• **patientId**: Integer representing the patient’s identification.

• **evaluationDate**: Evaluation’s date.

• **IdQuestion**: Integer representing the id of the question, is a foreign key for the table questions.

• **response**: Boolean containing the response to the question.

The risk’s levels are stored in the table evaluations, this table has the following columns:

• **patientId**: Integer representing the patient’s identification, foreign key for the table responses.

• **evaluationDate**: Evaluation’s date, foreign key for the table responses.

• **Suicide/Self-Aggress./Heteroagress/Escape/Pathology Risk**: Columns store the risk’s levels calculated by the decision trees.

• **Medic Suicide/Self-Aggress./Heteroagress/Escape/Pathology Risk**: Columns store the risk’s levels given by the health professionals when answering the questionnaire.
## Evaluation

### Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1 Risk Assessment Model’s Evaluation</td>
<td>55</td>
</tr>
<tr>
<td>5.2 Arisco system’s Evaluation</td>
<td>58</td>
</tr>
</tbody>
</table>
The system’s evaluation will focus in two different tasks. First, the performance of the new model composed by the decision tree classifiers is evaluated, described in section 5.1. Secondly the evaluation of the ARisCo system developed is made, a two month pilot test will be performed at the hospital collecting data from the patient’s evaluations and posterior a new system’s evaluation will be performed, described in section 5.2.

5.1 Risk Assessment Model’s Evaluation

This first stage of the evaluation was performed after creating the new risk assessment’s model composed by decision trees classifiers and the new questionnaire. An evaluation was made on the model to guarantee it offers good performance to be used on ARisCo system.

We will consider the metric accuracy to evaluate the performance of the classifiers, since it represents the percentage of instances correctly classified. The predictions made by the model are severity levels for mental health patients and the choice of which interventions will be based on these predictions. We must carefully take into account the classification errors the model can commit. Achieving a perfect accuracy is not always possible, but there are some errors that must be avoided.

We must be careful with the number of False Negatives instances concerning the risk’s classification, specially for the ‘Very High’ class. A patient that has the severity level of any risk as ‘Very High’ must be immediately taken care of. If these patients are classified as not being ‘Very High’ and do not receive the appropriate attention immediately dangerous outcomes are probable to occur.

We will control these cases using the metric sensitivity, since it measures the True Positive Rate: the entries correctly classified as positive. Our goal will be maximizing this metric concerning the class ‘Very High’, because having a high number of instances correctly classified as positive guarantees that we do not have many instances incorrectly classified as negative.

For the classifiers created we were able to obtain the values for the metrics of accuracy and sensitivity presented in table 5.1. These metrics were calculated using the confusion matrices for each risk category. Although the classifiers were developed using the balanced data samples, they were evaluated using the data sample collected from the hospital, composed by 321 evaluations from previous patients.

Confusion matrices regarding the decision trees classifiers developed: Suicide, Self-aggressiveness, Heteroaggressiveness and Escape are presented in the tables 5.2, 5.3, 5.4 and 5.5 respectively. These confusion matrices were created using the decision trees classifiers predictions, illustrated on the figures 5.1, 5.2, 5.3 and 5.4 respectively. On every decision tree’s final node, both the number of instances correctly classified and the miss classifications performed by the classifier are presented, being the numbers between parentheses respectively.

Observing table 5.1, it is possible to conclude that the tree’s classifiers offer better performance metrics
Table 5.1: Resulting metrics from the decision trees classifiers versus rules from hospital's current model

<table>
<thead>
<tr>
<th>Risk Category \ Metric</th>
<th>Number of instances</th>
<th>Accuracy</th>
<th>Sensitivity regarding 'Very High' class</th>
<th>Sensitivity regarding 'Very High' class</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Hospital's Rules</td>
<td>Decision Trees</td>
<td>Hospital's Rules</td>
</tr>
<tr>
<td>Suicide</td>
<td>59 VH/321</td>
<td>58.3%</td>
<td>83.49%</td>
<td>42.4%</td>
</tr>
<tr>
<td>Self-Aggress.</td>
<td>29 VH/321</td>
<td>93.5%</td>
<td>84.74%</td>
<td>79.3%</td>
</tr>
<tr>
<td>Heteroaggress.</td>
<td>33 VH/321</td>
<td>89.1%</td>
<td>94.70%</td>
<td>84.8%</td>
</tr>
<tr>
<td>Escape</td>
<td>14 VH/321</td>
<td>93.8%</td>
<td>97.20%</td>
<td>71.40%</td>
</tr>
</tbody>
</table>

Compared to the current rules of the hospital, the decision trees make a risk's prediction closer to the doctors and nurses predictions, the predictions on the data sample collected from the hospital. The only classifier's metric that is lower than the metric from the current hospital's rules is the accuracy regarding self-aggressiveness risk. The decision tree regarding this risk category has an accuracy of 84.74% while hospital's rules have 93.5%. But observing the values of sensitivity for this risk category, the tree classifier offers a much higher value than the hospital's rules. The trees classify 82.75% of the 'Very High' instances correctly while the hospital's rules only classify 79.3%.

Table 5.2: Confusion Matrix for Suicide Decision Tree

<table>
<thead>
<tr>
<th>Classified as (Decision Tree)</th>
<th>Actual Class (Data Sample)</th>
<th>Very High</th>
<th>not Very High</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Very High</td>
<td>43</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>not Very High</td>
<td>37</td>
<td>225</td>
</tr>
</tbody>
</table>

Observing table 5.2 and the decision tree classifier in figure 5.1 we can conclude that between the 321 instances being evaluated for the suicide risk, the tree classified correctly 268 instances (highlighted in green) and made 53 miss-classifications. 16 of these mistakes are worrying, because they were classified as being 'not Very High' when in fact they were 'Very High'. These miss-classifications are highlighted on the table 5.2 in red and also in the tree leaves where the miss-classifications occurs.

24 'Very High' instances are evaluated by the first right final node in the tree, when the patient presents the factors described in questions 2-“Maintenance of Suicidal Ideation” and 1.1-“High lethality of the act”. As concluded in the study performed at section 4.1.5.A, these were the questions that have the strongest relationship with the final risk category of suicide.
Observing table 5.3 and the decision tree classifier presented in figure 5.2 we can conclude that among the 321 instances being evaluated for the self-aggressiveness risk, the decision tree classified correctly 272 instances and made 49 miss-classifications. 5 instances were classified as being 'not Very High' when they were 'Very High', highlighted in red in the confusion matrix and also in the tree leaf where the classification errors occur. In figure 5.2, it is possible to observe that almost every 'not Very High' cases are classified in the first level of the tree and the 'Very High' cases on the second.

Observing table 5.4 and the tree classifier illustrated in figure 5.3 we can conclude that among the 321 instances being evaluated for the heteroaggressiveness risk, the tree classified correctly 304 instances
and made 17 mistakes. 4 of these miss-classifications are instances classified as being 'not Very High' when they were 'Very High', marked in red on the table, and in the tree leaves where the misclassification happens.

The factor that classifies almost every 'Very High' instances in the heteroaggressiveness category is question 17-“Current physical aggressiveness”, which was considered the question with the strongest relationship with the final heteroaggressiveness risk in the study performed at section 4.1.5.

With table 5.5 and the decision tree classifier illustrated at figure 5.4 it is concluded that among the 321 instances being evaluated for the escape risk, the tree classified correctly 312 instances and made 9 mistakes. 4 instances were classified as being 'not Very High' when in fact they were 'Very High', marked in red on the table, and in three leaves where the classification error occurs.

Almost every 'not Very High' instance is classified in the first final tree node with the questionnaire’s question 10 - "Abuse of alcohol or substances", in the study performed on section 4.1.5. This was the question that had the strongest relationship with the final level of risk for the escape category.

This is the decision tree classifier that presents the better metrics of all the classifiers developed.

5.2 Arisco system’s Evaluation

To evaluate the accuracy of the ARisCo system, composed by the decision tree classifiers described at section 4.3, the new questionnaire described at section 4.4 and the web based system developed
Figure 5.3: Heteroaggressiveness Tree Classification

Figure 5.4: Escape Tree Classification
described at section 4.5, a pilot test was performed at the hospital. This pilot test had the duration of 2
months and was performed by 3 nurses at Dª Estefânia Hospital.
During the pilot test, the nurses involved had the job of using the ARisCo system to perform the risk
assessment for the children hospitalized. The data collected by the system was stored to evaluate how
the ARisCo system performs in real life situations and how the risk assessment performed by the new
classifiers matches or not the risk assessment performed by the health professionals.

5.2.1 Data Sample’s Demographic Information

During the 2 month pilot test 55 evaluations were collected with the following characteristics:

- The evaluations collected refer to 27 patients.
- The maximum number of evaluations per patient is 4, having 5 patients hospitalized with 4 evalua-
tions.
- The minimum number of evaluation per patient is 1, having 14 patients hospitalized with 1 evalua-
tion.
- The maximum period of a patient hospitalization is 4 weeks.
- The minimum period of a patient hospitalization, not considering the patients with only 1 evaluation,
is 1 week.

Regarding the risk assessment performed by the decision tree classifiers, it was not displayed to the
health professionals, but their classifications were stored in order to compare the evaluations performed
by the decision trees with the ones performed by the health professionals.

The information regarding the classifications performed by the decision tree classifiers is presented at
table 5.6.

Table 5.6: Classification performed by the decision trees during the pilot test

<table>
<thead>
<tr>
<th>Risk Category</th>
<th>not Very High</th>
<th>Very High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suicide Risk</td>
<td>42</td>
<td>13</td>
</tr>
<tr>
<td>Self-Aggressiveness Risk</td>
<td>36</td>
<td>19</td>
</tr>
<tr>
<td>Heteroaggressiveness Risk</td>
<td>43</td>
<td>12</td>
</tr>
<tr>
<td>Escape Risk</td>
<td>54</td>
<td>1</td>
</tr>
<tr>
<td>Risk of Clinical Pathology</td>
<td>44</td>
<td>11</td>
</tr>
</tbody>
</table>

Besides the decision trees’ final classifications, the health professionals to classify the patients with
the risk’s level they considered appropriate. The information regarding the risk assessment performed
by the health professionals involved in the pilot test is presented at table 5.7.
Table 5.7: Classification performed by the health professionals during the pilot test

<table>
<thead>
<tr>
<th>Risk Category</th>
<th>not Very High (number of evaluations)</th>
<th>Very High (number of evaluations)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suicide Risk</td>
<td>47</td>
<td>8</td>
</tr>
<tr>
<td>Self-Aggressiveness Risk</td>
<td>49</td>
<td>6</td>
</tr>
<tr>
<td>Heteroaggressiveness Risk</td>
<td>46</td>
<td>9</td>
</tr>
<tr>
<td>Escape Risk</td>
<td>51</td>
<td>4</td>
</tr>
<tr>
<td>Risk of Clinical Pathology</td>
<td>51</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 5.8: Resulting metrics from the decision trees classifiers, Pilot Test

<table>
<thead>
<tr>
<th>Risk Category \ Metric</th>
<th>Numer of instances</th>
<th>Accuracy ARisCo</th>
<th>Sensitivity regarding 'Very High' class ARisCo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suicide Risk</td>
<td>8 VH/55</td>
<td>87.28%</td>
<td>87.5%</td>
</tr>
<tr>
<td>Self-Aggress. Risk</td>
<td>6 VH/55</td>
<td>72.73%</td>
<td>83.33%</td>
</tr>
<tr>
<td>Heteroaggress. Risk</td>
<td>9 VH/55</td>
<td>94.54%</td>
<td>100%</td>
</tr>
<tr>
<td>Escape Risk</td>
<td>4 VH/55</td>
<td>94.54%</td>
<td>25%</td>
</tr>
</tbody>
</table>

5.2.2 Risk Assessment Evaluation

To evaluate the risk assessment performed by the decision trees classifiers, the risk assessment performed by the health professionals on the pilot test was used. The following sections are dedicated to the evaluation of the risk assessment process made by the decision trees, comparing the final risks given by the trees to the ones given by the health professionals. To evaluate the classifiers, for each one of the risk categories the metrics of accuracy and the sensitivity regarding the 'Very High' class of risk were calculated. The decision trees performance metrics calculated are presented on table 5.8.

5.2.2.A Suicide Risk

Regarding the suicide risk, the risk assessment performed by the classifier and the one made by the health professionals was summarized in a confusion matrix presented in table 5.9. Observing table 5.8 it is possible to conclude that the ARisCo system presents a higher accuracy and sensitivity regarding the 'Very High' than the hospital current instrument. We started with a sensitivity regarding the 'Very High' from the current hospital rules of 42.4% and achieved a sensitivity of 87.5% with ARisCo system. Also the accuracy increased from 58.3% to 87.28%. The tree classifier only performed one error regarding the class 'Very High', highlighted at red on table 5.9.
5.2.2.B Self-aggressiveness Risk

Table 5.10: Confusion Matrix for Self-aggressiveness Risk, Pilot Test

<table>
<thead>
<tr>
<th></th>
<th>Classified as (Decision Tree)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Very High</td>
</tr>
<tr>
<td>Health Prof.</td>
<td>Very High</td>
</tr>
<tr>
<td>Classification</td>
<td>not Very High</td>
</tr>
</tbody>
</table>

Observing the performance metrics presented on table 5.8 it is possible to conclude that the current hospital’s instrument offers a higher value for accuracy than ARisCo, regarding the self-aggressiveness risk category.

Although the accuracy is not so positive as the hospital’s instrument, ARisCo system offers a better performance classifying the ‘Very high’ instances, having a sensitivity of 83.33% while the hospital rules present a sensitivity of 79.3%. The decision tree in the data sample collected performed one concerning mistake, highlighted in red on confusion matrix presented on 5.10.

5.2.2.C Heteroaggressiveness Risk

Observing the performance metrics presented at table 5.8, it is possible to conclude that the ARisCo system offers higher values for both metrics of accuracy and sensitivity regarding the risk of heteroaggressiveness. The ARisCo system presents a sensitivity of 100%, which means that the decision tree classifier developed classified all the ‘Very High’ cases equally to the evaluation made by the health professionals. Observing the confusion matrix on table 5.11, the classifier did not perform any concerning classification mistakes, having classified correctly all the ‘Very High’ cases. This is the risk category where ARisCo system presented better results.

Table 5.11: Performance Metrics for Heteroaggressiveness Risk, Pilot Test

<table>
<thead>
<tr>
<th></th>
<th>Classified as (Decision Tree)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Very High</td>
</tr>
<tr>
<td>Health Prof.</td>
<td>Very High</td>
</tr>
<tr>
<td>Classification</td>
<td>not Very High</td>
</tr>
</tbody>
</table>

5.2.2.D Escape Risk

Table 5.12: Performance Metrics for Escape Risk, Pilot Test

<table>
<thead>
<tr>
<th></th>
<th>Classified as (Decision Tree)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Very High</td>
</tr>
<tr>
<td>Health Prof.</td>
<td>Very High</td>
</tr>
<tr>
<td>Classification</td>
<td>not Very High</td>
</tr>
</tbody>
</table>
Observing the performance metrics presented on the table 5.8 it is concluded that ARisCo system offers a higher value for the metric of accuracy than the current hospital’s instrument on the risk category of escape, but it offers a very low metric of sensitivity regarding the ‘Very High’ cases. Observing the confusion matrix presented at table 5.12 the decision tree classifier only classified correctly 1 ‘Very High’ case in the 4 cases evaluated by the health professionals. ARisCo system performed 3 concerning mistakes, highlighted in red on the table 5.12. This is the lower metric value achieved by the system on the pilot test.

It is also possible to observe that all instances belonging to the 'not Very High' were all correctly classified.

### 5.2.2.E Risk of Clinical Pathology

Regarding the new risk added to the system, Risk of Clinical Pathology, it is not possible to make comparisons with the current hospital's instrument. Observing the confusion matrix presented on table 5.13, it is possible to calculate that the classifier created by the health professionals for ARisCo system offers an accuracy of 81.81% and a sensitivity regarding the 'Very High' class of 57.14%.

#### Table 5.13: Confusion Matrix for Risk of Clinic Pathology, Pilot Test

<table>
<thead>
<tr>
<th>Health Prof. Classification</th>
<th>Classified as (Decision Tree)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Very High</td>
<td>Very High</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>not Very High</td>
<td>not Very High</td>
<td>7</td>
<td>41</td>
</tr>
</tbody>
</table>

### 5.2.2.F Conclusions

From the study performed on the data collected during the pilot test it was possible to conclude that the trees performed well on a real life environment. It is important to understand that the amount of data collected in the two month testing is small and it is normal that the performance metrics are not so high as the ones from the initial data sample.

Even with a lower amount of data, the trees for the risks of Suicide, Self-Aggressiveness and Heteroaggressiveness all presented higher value for the metric of sensitivity than the performance metrics from the current hospital's instrument measured with the initial data sample. This means that the trees are efficient classifying the 'Very High' cases and produce less classification errors considered concerning, the ones where the patient is classified with a risk level lower than the one the health professionals consider appropriate.

For the suicide risk category it was possible to increase the metric of sensitivity to 87.5% which is more than the double of performance classifying 'Very High' instances. For the heteroaggressiveness risk category the performance metric of sensitivity was increased to 100%, the ARisCo system classified
correctly all the 'Very High' cases on the data sample.
The lower performance metrics were obtained on the Escape tree classifier, although with the ARisCo system the metric of accuracy was increased, the metric of sensitivity is lower, having a poor performance classifying the 'Very High' instances. It must be taken in consideration that even if the escape tree presents a sensitivity of 25 %, in the data sample there exist only 4 ‘Very High’ instances and the tree only missed classifying 3 of them.

5.2.3 Questionnaire’s Questions Evaluation

To understand the relations between the new questionnaire’s questions and the final risk’s levels present on the data sample collected during the pilot test, statistical studies were performed similar to the study performed at section 4.1.5.

5.2.3.A Correlation Coefficient

The equation of Pearson Coefficient [21] was used, similar to the study performed in section 4.1.5.A, the results from this study are displayed on table B.2. From this study it was possible to conclude:

- Regarding suicide risk, the questions that have stronger correlation are Question 2 - “Maintenance of Suicidal Ideation” and Question 1.1 - “Suicide attempt: Severe (egg hanging, drowning, firearm)”, having the correlation coefficients of 0.7933 and 0.6827 respectively.

- Regarding the self-aggressiveness risk category, it is possible to conclude that the question with the strongest relation with the final level of risk is Question 12 - “Self-sustaining behaviors maintained and recurrent”, having a correlation coefficient of 0.5663.

- On the Heteroaggressiveness risk category, it is possible to conclude that there are three questionnaire’s questions that present strong relation with the heteroaggressiveness level of risk: Question 16 - “Current physical aggressiveness” with a correlation coefficient of 0.6835, Question 14 - “Tense facial expression, loud voice and accelerated speech, stare and threatening or avoidance of eye contact with the observer; Tension / threatening posture” with a correlation coefficient of 0.6794.

- Regarding the risk of escape, the relations found are not so strong as the ones in the previous risks category. Question 16 - “Current physical aggressiveness” is the question with the strongest relation with the final risk level, with a correlation coefficient of 0.4126.

- Talking about the new risk category added to the questionnaire, Risk of Clinical Pathology, question 20 - “Risks associated with organic pathology (epilepsy, eating disorders, encephalitis)” is the
question added to calculate this risk category, present the correlation coefficients of 0.3563. This relationship was expected to be higher.

• Regarding the new questionnaire's questions:

  Question 3 - "Structured suicide plan (ex. Farewell letter, preparations)." presents a moderate association with the risk categories of Suicide and Auto-aggressiveness.

  Question 18-"Child does not verbally communicate." present a moderate association to the risk category of Pathology.

It was possible to identify questions that do not present relationships with any of the risk categories. The following questions presented correlation coefficients lower than 0.25, they all belong to the initial questionnaire: Question 4-"Family history of suicide.", Question 6- "Significant recent loss (egg mourning, material loss)," and Question 9- "History of child abuse".

5.2.3.B Logistic Regression

The same study performed on section 4.1.5.B was applied to the new data sample collected. A LR model was estimated for each one of the risk categories to understand the strongest and weakest relations between the questions and risk categories. Regarding the relationships found in the pilot's data sample, the strongest ones found are:

• The questions with the strongest associations with the suicide risk category are: Question 1.1- "Suicide attempt:Severe (egg hanging, drowning, firearm)". Question 2-"Maintenance of Suicidal Ideation" and Question 8 - "Feelings of hopelessness, lack of hope and low marked self-esteem."

• For the self-aggressiveness risk category the questions with the strongest association are: Question 12-"Self-sustaining behaviors maintained and recurrent" and Question 14-"Tense facial expression, loud voice and accelerated speech, stare and threatening or avoidance of eye contact with the observer; Tension / threatening posture".

• Regarding the heteroaggressiveness risk, the strongest associations were found in the questions: 14-"Tense facial expression, loud voice and accelerated speech, stare and threatening or avoidance of eye contact with the observer; Tension / threatening posture", 15-"Current verbal aggressiveness." and 16-"Current physical aggressiveness."

• In the escape risk category the question with the strongest association is Question 16-"Current physical aggressiveness."

---

1Study performed in collaboration with Professor Dina Salvador, School of Technology, Polytechnic Institute of Setúbal
• The new risk added showed a strong association with Question 19-“Risks associated with organic pathology (epilepsy, eating disorders, encephalitis).”

Using the LR model it was also possible to find which questions do not present associations with any risk category, being irrelevant for the risk assessment process:

• The new questionnaire’s questions: 3 - “Structured suicide plan (ex. Farewell letter, preparations).” and 18-“Child does not verbally communicate.” present very weak association with the risks.

• Regarding to the questions that belong to the initial questionnaire: Question 4-“Family history of suicide.”, Question 5-“Unprotected child (egg. Absence of reference figure, family / institutional).” and Question 9 -“History of child abuse”, also present weak association with the risks.

5.2.3.C Conclusions

From the both studies performed in sections 5.2.3.A and 5.2.3.B it is possible to conclude that the results regarding the questions with the strongest relationships match the ones from the initial correlation studies performed in the initial data sample collected, described at section 4.1.5.

For the risk of suicide, the question with the most impact on the level of risk are Question 1.1 - “Suicide attempt: Severe (egg hanging, drowning, firearm)” and Question 2 - “Maintenance of Suicidal Ideation”. For self-aggressiveness one of the questions that have more impact on the level of risk is Question 12 - “Self-sustaining behaviors maintained and recurrent”. For the heteroaggressiveness, also in both studies the question that affect most the level of risk for this category is Question 16 - “Current physical aggressiveness”.

Even with the modifications performed on the new questionnaire’s, there are still questions that do not present relevance for the prediction of the risks. Both studies, using the correlation coefficients and the LR model show that the following questions do not have impact in any of the risk categories:

• Questions belonging to the initial questionnaire Question 4-“Family history of suicide.”, Question 6- “Significant recent loss (egg mourning, material loss).”,Question 9 -“History of child abuse”. These questions on the initial data sample presented a moderate/weak association with the risks’, as shown on table B.1 and they still are not relevant to evaluate the risk.

• Regarding to the new added questionnaire’s questions: Question 18-“Child does not verbally communicate.” and 3 - “Structured suicide plan (ex. Farewell letter, preparations).” do not present any association with any risk category using the LR models approach but present moderate associations using the correlation coefficients.
5.2.4 Pilot Test Data sample’s Decision Trees

Decision trees classifiers were developed using the data sample collected on the pilot test with the goal of validating the initial classifiers. Using the data sample composed by 55 evaluations, the C4.5 algorithm [18] was used to develop decision trees. The goal of this study is for the decision trees created from the pilot test evaluations be similar to the ones created using the initial data sample collected, meaning that the ARisCo’s classifiers are suitable to perform the risk assessment. This study also allows to understand which questionnaire’s questions are relevant for each risk category assessment, the set of questions used in the decision tree are the most relevant for predicting the risk. The data sample was balanced for each risk category because number of instances belonging to the class ‘Very High’ was low. Cross-validation was used with the number of 10 folds. The pruning techniques of minimal number of instance per leaf and confidence factor were used to achieve a better performance.

5.2.4.A Suicide Tree

The tree classifier that presents the best performance for the pilot test data sample regarding the suicide risk category, is composed by only one node as presented at figure 5.5. This tree classifier offers an accuracy of 96.84% and a sensitivity of 97.9%. The classifier offers a very positive performance, classifying correctly 96.84% of the evaluations, using only Question 3 - “Structured suicide plan (ex. Farewell letter, preparations),”, one of the new questions added to the questionnaire.

![Figure 5.5: Suicide Tree Pilot Test](image)

5.2.4.B Self-aggressiveness Tree

The self-aggressiveness decision tree developed using the pilot test’s data sample is equal to the decision tree developed using the initial data sample, present in figure 4.2. This decision tree classifier presented an accuracy of 95.88% and a sensitivity of 95.8%.
5.2.4.C  Heteroaggressiveness Tree

The decision tree developed for the heteroaggressiveness risk category is presented at figure 5.6. This decision tree is similar to the one developed using the initial data sample, both questions 15 - "Current verbal aggressiveness." and 16 - "Current physical aggressiveness" are used to perform the risk assessment for this category. This decision tree classifier offers an accuracy of 96.7% and a sensitivity of 95.6%.

Figure 5.6: Heteroaggressiveness Tree Pilot Test

5.2.4.D  Escape Tree

The decision tree created for this risk category, illustrated in figure 5.7, is similar to the tree classifier created using the initial data sample. Both tree classifiers use the questions 17 - "Low adherence in therapeutic design (child / family)." and 7 - "Serious Socialization Problems / Social Isolation." to perform the risk assessment. The new decision tree offers an accuracy of 94.17% and a sensitivity of 96.2%.

5.2.4.E  Conclusions

From the decision trees' development, using the data sample collected from the pilot test, we concluded:

• Contrary to the study performed using the LR approach described at section 5.2.3.B, the new questionnaire's question 3 - "Structured suicide plan (ex. Farewell letter, preparations)." is important to perform the risk assessment for suicide, classifying correctly 96.84% of the data sample.

• The decision tree for the risk category of self-aggressiveness match the one developed from the initial data sample collected, it means that the risk assessment for this category in both data
samples is the same. This strongly supports that the decision tree offers a correct classification of the risk.

• Regarding the heteroaggressiveness classifier, the tree developed using the pilot test data sample is similar to the one developed using the initial data sample. Both classifiers predict the risk using the questions 15 - "Current verbal aggressiveness." and 16 - "Current physical aggressiveness". Contrary to the initial decision tree, the new classifier do not use question 14 - "Tense facial expression, loud voice and accelerated speech, stare and threatening or avoidance of eye contact with the observer; Tension / threatening posture".

• For the escape risk categories, both tree classifiers use the questions 17 - "Low adherence in therapeutic design (child / family)." and 7 - "Serious Socialization Problems / Social Isolation." to perform the risk assessment, but the new classifier developed do not use question 10 - "Abuse of alcohol or substances." contrary to the one initial developed.

• With this study and the ones performed in the previous section 5.2.3, we concluded that the questionnaire still have questions that are not relevant to predict the risk. Question 4 - "Family history of suicide.", Question 6 - "Significant recent loss (egg mourning, material loss) and "Question 9 - "History of child abuse" do not present associations with any risk category, as concluded in the correlation coefficient and LR studies performed. The questions also are not used in the decision trees. In the initial data sample they also present weak associations. These facts led us to conclude that these questions are not relevant for the risk assessment.

• Finally, we concluded that the model created for the ARisCo system offers a good classification of the risk because the classifiers developed based on the initial data sample collected are similar,
in the self-aggressiveness risk category equal, to the ones developed using the new data sample collected from the pilot test.

Several studies were performed using different approaches as described on sections 5.2.3 and 5.2.4, to validate the model developed composed by the questionnaire and the decision trees classifiers. From the three different approaches it was possible to reach the same conclusion regarding the questionnaire: Question 4 - "Family history of suicide.", Question 6 - "Significant recent loss (egg mourning, material loss)." and Question 9 - "History of child abuse" do not present any association with any risk category, which means these questions are not relevant to the risk’s prediction. Similar studies were performed on the initial data sample described section 4.1.5, these questions already presented low associations with the risk’s categories. For these reasons, the questions will be discussed with the health professionals involved in the project, to be removed from the questionnaire.
Conclusion

Contents

6.1 Contributions ............................................................. 72
6.2 Main Results ............................................................... 72
6.3 Conclusions ............................................................... 74
6.4 System Limitations and Future Work ................................. 75
Performing an efficient risk assessment is the key to offer good care to the hospitalized patients who suffer from mental diseases. The assessment of the risk is a task that must be performed very carefully. Doctors and nurses decide which treatment plans or interventions are appropriated for the patients based on that assessment. Using a system that has the capability of performing this process automatically can bring enormous improvements, not only for the life of hospitalized patients but also to health professionals. Motivated by improving the current situation of the area of child and adolescents mental health of Dª Estefânia Hospital the development of a new recommendation system, ARisCo, responsible for the risk assessment process was proposed.

6.1 Contributions

This project contributed to the area of mental health of children and adolescents of Dª. Estefânia Hospital by improving the current process of risk assessment. By replacing the current pen & paper instrument by the recommendation system ARisCo, the time consumed to perform an evaluation was decreased, specially by having decision trees classifiers that perform the risk assessment process automatically. It also helped decreasing the miss classifications errors and store patient’s data efficiently. Besides the main contribution for Dª. Estefânia Hospital, the ARisCo project presented in form of a poster in the events: XVII Symposium of the Portuguese Society of Suicidology in Tomar and on the event Journey of Child and Adolescent Mental Health Care on Dª. Estefânia Hospital. The poster displayed and presented in these events is illustrated on figure A.5 on attachments.

As discussed in the project’s presentations in these events, this project brings an all new perspective to the health care area. Joining engineering techniques and algorithms with the knowledge of the health professionals, it was possible to achieve conclusions and findings in the health area that would not have been possible to be achieve by only one of the parts involved.

6.2 Main Results

The first step of ARisCo system's development was the creation of a new classification model for the risk assessment process. An analysis was made on the current hospital's instrument. To perform this analysis, evaluations from previous hospitalized patients were collected and analyzed. From the analyses on the data sample collected it was concluded that the current hospital's instrument was not suitable for performing the risk assessment process. The current instrument is time consuming, the risk assessment was calculated by hand using rules, those rules were contradictory and the severity of risk’s levels attributed to the patient did not match the knowledge and opinion of the health professionals.
This situation led to a manipulation of the instrument by the health professionals, to obtain the risk's level they thought was suitable and not the one actually returned by the instrument.

Only 83.7% of the evaluations from the data sample, match the evaluation performed by the current rules. This means that on 16.3% of the evaluations the health professionals ignored the current instrument's risk assessment and performed their own. Regarding the suicide risk category only 58.3% of the evaluations the current instrument match the evaluation performed by the health professionals.

A new classification model was developed with the goal of performing a risk assessment closer to the one on the data sample collected. Classifiers in form of decision trees were created and all offer performance metrics superior to the ones from the current hospital's instrument, which means that the risk assessment performed by the trees classifier is closer to the one performed by the health professionals than the one performed by the rules from the current instrument.

The new tree classifiers offer accuracy values superior to 83% for every risk category. With the new model the accuracy for the suicide risk category increased from 58.3% to 83.49%. The values for the metric sensitivity regarding the 'Very High' class, in every risk category, also increased, having the average of 78.74%, which means that 78.74% of the evaluations collected the decision trees classify the risk's level as 'Very High' so as the health professionals do, while the current rules offer an average of sensitivity of 69.48%.

Regarding the current questionnaire it was concluded that it was inappropriate. The health professionals collaborating in the project scheduled several focus groups performed, where the current hospital's instrument was discussed among several doctors and nurses. It was found that the current questionnaire has irrelevant, ambiguous and duplicated questions and the medical terms were not updated. Several meetings were scheduled to create a new efficient, small and correct questionnaire for ARisCo.

After having the new model, composed by the decision tree classifiers and by the new questionnaire, ARisCo system was developed to perform the risk assessment process automatically. Having the system developed, a two month pilot test was performed in Dª. Estefânia Hospital.

On the pilot test a data sample composed by 55 evaluations was collected. In the few evaluations collected, the decision trees performed a risk assessment close to the health professionals having an average accuracy of 87.3%. In all the risk categories the accuracy increased comparing with the hospital's rules classifier using the initial data sample, except for the self-aggressiveness risk which we started with an accuracy of 93.5% and the ARisCo system present a value of 72.73%. Although the accuracy decreased, the tree classifier produces a positive value for the metric of sensitivity which means that it performs well while classifying the 'Very High' instances. The ARisCo offers a sensitivity of 83.33% for the self-aggressiveness risk while the hospital's instrument had 79.33%. The decision tree that presented a better performance was the Heteroaggressiveness risk tree, which offers an accuracy of 94.54% and a sensitivity of 100%, which means that almost every instances classified match the classi-
fication performed by the health professionals, classifying correctly all the 'Very High' instances.

Regarding the Escape risk category, the results were not so positive regarding the sensitivity metric for the 'Very High' instances, in four 'Very High' instances only one was well classified, having the value of 25% for the metric sensitivity. This value is low but it is important to have in consideration that the amount of data is low and there are few 'Very High' instances to classify in the data. We cannot take the conclusion that the classifier is not adequate by only evaluating 4 'Very High' instances.

Correlation and statistical studies were performed on the pilot test's data sample to validate the new questionnaire and also the decision tree classifiers. From these studies it was concluded that the questionnaire, even after its alteration, still presents questions that do not present any association with any risk category, being irrelevant to the risk assessment.

6.3 Conclusions

The first task performed in this project was a research on the project's scope: risk assessment in children and adolescents mental health care. It was concluded that there exist several studies and tools to perform the risk assessment for patients that suffer from mental diseases, but few focus on children and adolescents. GRiST was the only system found that offers a complete and detailed risk assessment focused on children and adolescents. Although children and adolescents mental health does not receive much attention it is extremely important because a lot of mental diseases can be more successfully treated in an early intervention and avoid lifelong mental diseases in the patient's future.

The development of ARisCo system consisted on three main phases, in each of them important conclusions were reached that had an impact on the following phase.

- The first main phase the analysis of the data sample composed by evaluations from previous hospitalized children and adolescents was made. It was concluded that the current hospital's instrument used to perform the risk assessment was not adequate, which led to a manipulation and contempt on the use of the instrument. With the several studies performed on the data sample and with the collaboration of the health professionals, it was concluded that the questionnaire had irrelevant, duplicate and ambiguous questions and the rules used to calculate the risk were also contradictory and confusing.

- The second phase consisted on the development of the new model to perform the risk assessment. We chose decision trees classifiers to perform the risk's classification whiting other possible approaches, because they are easy to be understood by the health professionals. The goal of this phase was to create a classification model that produces a risk assessment similar to the health professionals assessment, but instead of the hospital's instrument rules, we used...
decision trees and reduced the number of input questions. This classification model was evaluated using performance metrics, comparing its predictions to the ones in the data sample. It was concluded that the new model offers a risk assessment closer to the health professionals’ intuition and knowledge than the current hospital’s instrument. Besides the better performance, the model offers a simpler, faster and more intuitive way of classifying the risks.

• The third and last phase of the project consisted on the development of the ARisCo system to perform the risk assessment automatically. The system reduces the time consumed and reduces the probability of performing errors since in the current instrument, health professionals perform all the process manually. The ARisCo system was subjected a two month pilot test in the hospital, where the nurses involved in the test performed the patients’ evaluations. From the pilot test we concluded that ARisCo system offers a positive performance classifying the risk, producing high values for the metrics of accuracy and sensitivity. Although in some cases the metric of sensitivity was lower than expected, it is important to understand that the number of instances collected is low. It is not possible to conclude that a classifier is not suitable with the amount of instances collected in the pilot test.

It was concluded that ARisCo performed well in the pilot test and the feedback received from the nurses involved in the test period was positive.

All the proposed objectives for this project were successfully achieved. An efficient and appropriate recommendation system was developed to facilitate the process of risk assessment of the children and adolescents suffering from mental diseases hospitalized in D. Estefânia hospital. The main goal from the start, was helping the health professionals performing their work and so improve the conditions, treatment and life of the children and adolescents.

6.4 System Limitations and Future Work

Although all the project’s goals were achieved, there exists space to improve.

The ARisCo system is still being used in the hospital collecting data. The time available to produce this project was enough to collect patient’s evaluations to validate correctly the classification model developed. When there are enough evaluations stored, a new evaluation of the ARisCo should be performed to guarantee that the classification model produces positive performance metrics. Currently, the ARisCo system uses a questionnaire that the health professionals fill in and performs a risk assessment for the patient. The type of supervision or interventions the patient receives is based on the risk assessment performed. It is important to understand if the patient after being classified with a certain level of risk had a positive outcome. Instead of just take into consideration the assessment made by the health professionals to evaluate the correctness of the system, it would be interesting to
understand, based on the risks given by the ARisCo system and analyzing the process of each patient, if the chosen level of risk was adequate. By joining this information to ARisCo system, it would be possible to understand how the patients respond to the treatments and interventions made based on the system’s evaluation.

It would also be interesting to perform a probabilistic research on how the patients evolve within the risk level assigned and the symptoms and factors they present. Studying the data samples collected, it would be possible to predict how a patient with a certain level of risk and factors would progress, if the probability of that patient with those symptoms/factors is to maintain the level of risk, increase or decrease it.
Bibliography


[34] M. O’Rourke, G. Bailes, and J. Davies, Risk Assessment and Management. The British Psychological Society, November 2006.
Figures
**Figure A.1:** Final questionnaire Portuguese version

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Sim</th>
<th>Não</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Tentativa de suicídio.</td>
<td>1.1. Grave (exs. enforcamento, afogamento, arma de fogo)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.2. Recorrente</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Plano de suicidioestruturado (exs. cartaz despedida, preparativos).</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. História familiar de suicídio.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Criança desprotegida (exs. ausência de figura de referência, familiar / institucional).</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Perda recente marcante (exs. luto, perda material).</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Problemas de socialização graves/ isolamento social.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Sentimentos de desespero, falta de esperança e baixa autoestima marcada.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Abuso de álcool ou substâncias.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Sexo Masculino.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Comportamentos autolesivos mantidos e recorrentes.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Alteração da percepção, pensamento, comportamento ou deficiência mental.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. Expressão facial tensa, tom de voz alto e discurso acelerado, olhar fixo e ameaçador ou evitamento de contato ocular com o observador; Postura tensa/ameaçadora.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. Agressividade verbal atual.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. Baixa adesão no projeto terapêutico (criança/familia).</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18. Criança não comunica verbalmente.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19. Riscos associados a patologia orgânica (epilepsia, perturbações de comportamento alimentar, encefalite).</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Figure A.2: IRIS Portuguese version**

**SOCIODEMOGRAFIA - Ponderação 1**
- **Sexo**  
  Masculino → 1  
  Feminino → 0
- **Idade**  
  ≥ 45 → 1  
  < 45 → 0
- **Religiosidade**  
  Não → 1  
  Sim → 0

Existem factores de natureza religiosa ou espiritual suscetíveis de frenar a passagem ao acto?

**CONTEXTOS - Ponderação 2**  
Não → 0  
Sim → 2
- **Isolamento** - vive só, sem apoio familiar ou social?
- **Perda recente marcante** - luto, desemprego, perda material ou de estatuto
- **Doença física** - incapacitante ou terminal
- **Abuso actual** de álcool ou substâncias
- **Doença psiquiátrica grave** - descompensação actual de psicose, depressão maior unipolar ou bipolar, perturbação grave da personalidade
- **História de internamento psiquiátrico**
- **História familiar de suicídio**

**ESFERA SUICIDA**
- **História pessoal de comportamentos suicidários**  
  Ponderação 3  
  Não → 0  
  Sim → 3

  Considerar Sim em caso de 2 ou mais comportamentos **prévios** ou apenas 1 se grave  
  (método violento ou tendo justificado cuidados intensivos)

- **Plano suicida**
  Apura-se a existência de plano organizado, consistente, letal e exequível?
  - valorizar actos preparatórios recentes (exs: carta de despedida, testamento), bem como o acesso a meios letais (exs: arma de fogo, pesticidas / herbicidas)
  Não → 0  
  Sim → Atribuir **directamente** o valor 20 ao Score Total do Índice

82
**Figure A.3:** Hospital's current questionnaire Portuguese version

**Identificação de Fatores de Risco Clínico**

<table>
<thead>
<tr>
<th>Risco Suicidário</th>
<th>/ /</th>
<th>/ /</th>
<th>/ /</th>
<th>/ /</th>
<th>/ /</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Tentativa de Suicídio</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disruptiva (impulsiva com método grave)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recorrente</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Manutenção de Ideia Suicida</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Elevada letalidade do ato (método agressivo e/ou planeado; intenção de morte expressa ou inferida)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Perturbação do humor, ou elevada impulsividade, psicose.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Disfunção familiar grave e/ou história familiar de t. Suicídio ou suicídio</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Perdas afetivas importantes por morte ou término de relação de suporte afetivo fundamental</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Problemas de socialização graves/isolamento social</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Sentimentos de desespero, falta de esperança e baixa de autoestima marcada</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. História de abuso do menor; consumo de substâncias</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Sexo masculino</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Risco (MA/M/B)**

<table>
<thead>
<tr>
<th>Autogressividade</th>
<th>/ /</th>
<th>/ /</th>
<th>/ /</th>
<th>/ /</th>
<th>/ /</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. História de comportamentos de auto agressão mantidos e recorrentes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Comportamentos de autoagressão e existência de patologia psicótica, da personalidade ou debilidade mental</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Comportamentos de autoagressão com risco de suicídio</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Risco (MA/M/B)**

<table>
<thead>
<tr>
<th>Violência</th>
<th>/ /</th>
<th>/ /</th>
<th>/ /</th>
<th>/ /</th>
<th>/ /</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Expressão facial tensa, tom de voz alto e rápido, olhar fixo e ameaçador ou evitamento do contato ocular com observador</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Postura tensa/ameaçadora (punhos, dentes cerrados)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Agitação motora excessiva/movimentos agressivos (pontapés, muros a objetos)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Agressividade e/ou ameaças verbais: (sarcasmo, ridicularização, referencias menos abonatórias às diferenças, comentários de menosprezo, desconfiança, desafio)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Contestação e quebra frequente e mantida das regras do Serviço</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Percepções delirantes de cariz paranoide</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Ameaça de agressividade física dirigida a objetos ou a pessoas</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Agressividade expressa, física, dirigida a objetos ou a pessoas</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Risco (MA/M/B)**

<table>
<thead>
<tr>
<th>Risco de Fuga</th>
<th>/ /</th>
<th>/ /</th>
<th>/ /</th>
<th>/ /</th>
<th>/ /</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Pouca consistência na adesão da criança e/ou família ao projeto terapêutico</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. História prévia de “Fugas”</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. História prévia de comportamentos aditivos.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Risco (MA/M/B)**
Figure A.4: ARisCo system questionnaire’s page

ARisCo

Sistema de Recomendação para Análise de Risco em Saúde Mental de Criança e do Adolescente

Código de identificação do atente: (22122)

Identificação de fatores para o cálculo dos riscos: Suicidio, Autoagressividade, Heteroagressividade, Fuga e Patologia Clínica.

1. Ideações de intenção suicida.
   Sim: Não

2. Planeio de suicídio estruturado (ex. carta de despedida, preparatória).
   Sim: Não

3. Perda recente marcante (morte, luto, perda material).
   Sim: Não

4. Problemas de socialização graves ou tratamento social.
   Sim: Não

5. Sentimentos de desespero, falta de esperança e baixa autoestima marcada.
   Sim: Não

6. Comportamentos autoagressivos mandatórios e recorrentes.
   Sim: Não

7. Alteração da percepção, pensamento, comportamento ou deficiência mental.
   Sim: Não

8. Expressão facial tenso, tom de voz alto e discurso acelerado, olhar fixo e ameaçador ou esvaziamento de contato ocular com o observador, Postura tensa/ameaçadora.
   Sim: Não

   Sim: Não

10. Agressividade física atual.
    Sim: Não

    Sim: Não

12. Criança não comunica verbalmente.
    Sim: Não

13. Riscos associados à patologia orgânica (Epilepsia, perturbações de comportamento alimentar, enxaqueca).
    Sim: Não

Níveis de risco a atribuir ao cliente:

Selecione o nível de risco que considera o mais adequado a atribuir ao cliente.

Risco Suicídio: Muito Alto - não Muito Alto
Risco Autoagressividade: Muito Alto - não Muito Alto
Risco Heteroagressividade: Muito Alto - não Muito Alto
Risco de Fuga: Muito Alto - não Muito Alto
Risco de Patologia Clínica: Muito Alto - não Muito Alto

Submeter
**Figure A.5: Arisco System Poster**

**Arisco: Sistema de Recomendação para Análise de Risco em Saúde Mental da Criança e do Adolescente**

**Contexto**

O projeto Arisco consiste no desenvolvimento de um sistema de recomendação responsável pela realização da avaliação do risco clínico dos pacientes da área de Pedopsiquiatria do Centro Hospitalar Lisboa Central-Hospital de Estefânia, melhorando o processo de avaliação do risco atual.

**Objetivos**

- Analisar o sistema de avaliação do risco clínico na área de Pedopsiquiatria, do Centro Hospitalar Lisboa Central-Hospital de Estefânia;
- Criar um sistema de apoio à decisão para a avaliação do risco clínico em Pedopsiquiatria, nomeadamente para os riscos de Suicídio, Auto e Heteroagressividade e Fuga.

**Metodologia**

Para o desenvolvimento deste sistema está a ser efetuado um estudo estatístico utilizando o instrumento da avaliação do risco existente no hospital. Para este estudo serão utilizadas 321 avaliações de pacientes hospitalizados entre os anos 2015 e 2017. Este instrumento utilizado atualmente consiste numa lista de perguntas relacionadas aos riscos de Suicídio, Autoagressividade, Heteroagressividade e Fuga. Ao responder e as perguntas existem regras responsáveis por calcular qual o risco a atribuir com base nas respostas dadas pelos profissionais de saúde.

**Identificação de Fatores de Risco Clínico**

<table>
<thead>
<tr>
<th>Fatores de Risco</th>
<th>Peso</th>
<th>Nota</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Tensões de Suicídio</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Transtornos de Identidade Sexual</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Exame psicológico ou neurológico</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Intenção de morte, tentativa de suicídio, infecciosidade, angústia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Exame psicológico ou neurológico</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Fatores importantes por morte ou terminação de relação de supervivência</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Fatores de suicídio</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figura 1: Perguntas do instrumento atual referentes ao Risco de Suicídio**

**Resultados**

Após o estudo realizado foi possível identificar que os graus dos riscos desvios pelo instrumento utilizado atualmente para classificar o risco, não coincidem com a avaliação dos profissionais, incluindo médicos e enfermeiros. Fazendo com que o instrumento atual seja manipulado de maneira a serem obtidos os graus das avaliações desejados e não os que são retornados na realidade pelos regras do instrumento.

As arcos de decisão criadas a partir dos dados recobertos conseguem atribuir uma avaliação do risco semelhante a avaliação realizada pela equipa médica.

**Figura 2: Arco de decisão do risco de Suicídio**

Foi também possível detectar algumas perguntas que não são necessárias para realizar a avaliação do risco. Ao eliminar estas perguntas reduziu-se o tamanho do questionário de 25 perguntas para 16, mantendo as métricas de precisão.

**Figura 3: Perguntas do instrumento atual referentes ao Risco de Fuga**

**Conclusões**

Com o estudo retrospectivo a criação das novas arcos de decisão conclui-se que o questionário utilizado para a avaliação do risco clínico não está de acordo com a intenção da equipa médica, fazendo com que não seja utilizado sistematicamente ou que seja manipulado de maneira a obter o grau do risco que a equipa médica possa ser o mais adequado.

As arcos de decisão criadas para cada um dos riscos apresentam uma precisão superior ao questionário utilizado no hospital e para a avaliação de equipa médica. Para além de uma maior precisão, apresentam menos perguntas e oferecem uma melhor compreensão no processo de cálculo do risco. No futuro este instrumento irá ser criado em forma de uma plataforma informática e testado.
Tables
Table B.1: Correlation coefficients between questionnaire’s questions and final risk’s levels - Hospital’s data sample

<table>
<thead>
<tr>
<th></th>
<th>Suicide Risk</th>
<th>Self-Aggressiveness Risk</th>
<th>Heteroaggressiveness Risk</th>
<th>Escape Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sui. 1.1</td>
<td>0.368131</td>
<td>0.177444</td>
<td>-0.1581</td>
<td>-0.00549</td>
</tr>
<tr>
<td>Sui. 1.2</td>
<td>0.452487</td>
<td>0.112682</td>
<td>-0.13287</td>
<td>-0.10064</td>
</tr>
<tr>
<td>Sui. 2</td>
<td>0.604877</td>
<td>0.213063</td>
<td>-0.17357</td>
<td>-0.07086</td>
</tr>
<tr>
<td>Sui. 3</td>
<td>0.607739</td>
<td>0.013771</td>
<td>-0.16161</td>
<td>-0.00861</td>
</tr>
<tr>
<td>Sui. 4</td>
<td>0.273566</td>
<td>0.166071</td>
<td>0.134032</td>
<td>0.176929</td>
</tr>
<tr>
<td>Sui. 5</td>
<td>0.311306</td>
<td>0.085784</td>
<td>-0.00419</td>
<td>-0.11227</td>
</tr>
<tr>
<td>Sui. 6</td>
<td>0.248401</td>
<td>0.065271</td>
<td>-0.02304</td>
<td>0.013857</td>
</tr>
<tr>
<td>Sui. 7</td>
<td>0.179315</td>
<td>0.142778</td>
<td>0.047629</td>
<td>0.013857</td>
</tr>
<tr>
<td>Sui. 8</td>
<td>0.396722</td>
<td>0.122436</td>
<td>-0.17977</td>
<td>-0.08738</td>
</tr>
<tr>
<td>Sui. 9</td>
<td>0.182891</td>
<td>0.021882</td>
<td>0.084087</td>
<td>-0.17923</td>
</tr>
<tr>
<td>Sui. 10</td>
<td>-0.08086</td>
<td>-0.03579</td>
<td>0.2235</td>
<td>0.206036</td>
</tr>
<tr>
<td>Self. 1</td>
<td>0.275049</td>
<td>0.33105</td>
<td>-0.00693</td>
<td>-0.03333</td>
</tr>
<tr>
<td>Self. 2</td>
<td>0.010621</td>
<td>0.376482</td>
<td>0.159203</td>
<td>0.093352</td>
</tr>
<tr>
<td>Self. 3</td>
<td>0.370268</td>
<td>0.463912</td>
<td>-0.05116</td>
<td>-0.0516</td>
</tr>
<tr>
<td>Heter. 1</td>
<td>-0.09402</td>
<td>0.130125</td>
<td>0.520061</td>
<td>0.161551</td>
</tr>
<tr>
<td>Heter. 2</td>
<td>-0.11364</td>
<td>0.065155</td>
<td>0.414629</td>
<td>0.319845</td>
</tr>
<tr>
<td>Heter. 3</td>
<td>-0.13182</td>
<td>0.112917</td>
<td>0.577119</td>
<td>0.237989</td>
</tr>
<tr>
<td>Heter. 4</td>
<td>-0.10678</td>
<td>-0.06814</td>
<td>0.49065</td>
<td>0.252448</td>
</tr>
<tr>
<td>Heter. 5</td>
<td>-0.0905</td>
<td>-0.00525</td>
<td>0.529659</td>
<td>0.26231</td>
</tr>
<tr>
<td>Heter. 6</td>
<td>-0.08691</td>
<td>-0.00221</td>
<td>0.342088</td>
<td>0.267705</td>
</tr>
<tr>
<td>Heter. 7</td>
<td>-0.11939</td>
<td>0.095541</td>
<td>0.648595</td>
<td>0.257107</td>
</tr>
<tr>
<td>Heter. 8</td>
<td>-0.1359</td>
<td>0.072675</td>
<td>0.816742</td>
<td>0.223728</td>
</tr>
<tr>
<td>Esc. 1</td>
<td>-0.02791</td>
<td>0.060457</td>
<td>0.245857</td>
<td>0.324317</td>
</tr>
<tr>
<td>Esc. 2</td>
<td>-0.09076</td>
<td>-0.01142</td>
<td>0.363477</td>
<td>0.412893</td>
</tr>
<tr>
<td>Esc. 3</td>
<td>-0.01078</td>
<td>-0.08818</td>
<td>0.17653</td>
<td>0.519155</td>
</tr>
<tr>
<td>Question</td>
<td>Suicide Risk</td>
<td>Self-Aggressiveness Risk</td>
<td>Heteroaggressiveness Risk</td>
<td>Escape Risk</td>
</tr>
<tr>
<td>----------</td>
<td>--------------</td>
<td>--------------------------</td>
<td>--------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>1.1</td>
<td>0.682745</td>
<td>0.064626</td>
<td>-0.15478</td>
<td>-0.098</td>
</tr>
<tr>
<td>1.2</td>
<td>0.470848</td>
<td>0.24358</td>
<td>-0.08592</td>
<td>-0.0544</td>
</tr>
<tr>
<td>2</td>
<td>0.793323</td>
<td>0.475811</td>
<td>-0.19565</td>
<td>0.065379</td>
</tr>
<tr>
<td>3</td>
<td>0.329843</td>
<td>0.388889</td>
<td>-0.06019</td>
<td>-0.03811</td>
</tr>
<tr>
<td>4</td>
<td>0.083046</td>
<td>-0.098</td>
<td>-0.12388</td>
<td>0.191176</td>
</tr>
<tr>
<td>5</td>
<td>-0.30198</td>
<td>-0.1103</td>
<td>0.253821</td>
<td>0.145144</td>
</tr>
<tr>
<td>6</td>
<td>-0.06078</td>
<td>0.137464</td>
<td>0.046337</td>
<td>-0.13202</td>
</tr>
<tr>
<td>7</td>
<td>0.275949</td>
<td>0.234051</td>
<td>0.1895</td>
<td>0.035811</td>
</tr>
<tr>
<td>8</td>
<td>0.46889</td>
<td>0.1625</td>
<td>0.007207</td>
<td>0.035936</td>
</tr>
<tr>
<td>9</td>
<td>-0.11554</td>
<td>-0.098</td>
<td>0.065379</td>
<td>-0.07843</td>
</tr>
<tr>
<td>10</td>
<td>-0.14437</td>
<td>0.064826</td>
<td>0.002866</td>
<td>0.351164</td>
</tr>
<tr>
<td>11</td>
<td>-0.12269</td>
<td>-0.07059</td>
<td>0.305654</td>
<td>0.157806</td>
</tr>
<tr>
<td>12</td>
<td>0.306664</td>
<td>0.566339</td>
<td>-0.16892</td>
<td>0.103129</td>
</tr>
<tr>
<td>13</td>
<td>-0.09308</td>
<td>-0.04365</td>
<td>0.36132</td>
<td>0.191096</td>
</tr>
<tr>
<td>14</td>
<td>0.013242</td>
<td>0.35442</td>
<td>0.679369</td>
<td>0.173787</td>
</tr>
<tr>
<td>15</td>
<td>-0.09308</td>
<td>0.238773</td>
<td>0.599314</td>
<td>0.191096</td>
</tr>
<tr>
<td>16</td>
<td>-0.06078</td>
<td>0.288675</td>
<td>0.683464</td>
<td>0.412561</td>
</tr>
<tr>
<td>17</td>
<td>-0.03766</td>
<td>0.085184</td>
<td>-0.18843</td>
<td>0.255655</td>
</tr>
<tr>
<td>18</td>
<td>-0.05614</td>
<td>-0.04762</td>
<td>-0.06019</td>
<td>-0.03811</td>
</tr>
<tr>
<td>19</td>
<td>-0.20628</td>
<td>-0.17496</td>
<td>-0.09829</td>
<td>-0.14003</td>
</tr>
</tbody>
</table>
Code
Algorithm C.1: Suicide tree JSON

```json
suicideTree = {
    "question": "Quest_3",
    "FinalNode": false,
    "responseNo": {
        "FinalNode": true,
        "response": "nao Muito Alto",
        "id": "suicide_final_1"
    },
    "responseYes": {
        "FinalNode": false,
        "question": "suicideCounter_3",
        "responseNo": {
            "FinalNode": true,
            "response": "nao Muito Alto",
            "id": "suicide_final_2"
        },
        "responseYes": {
            "FinalNode": false,
            "question": "Quest_12",
            "responseNo": {
                "FinalNode": true,
                "response": "Muito Alto",
                "id": "suicide_final_3"
            },
            "responseYes": {
                "FinalNode": true,
                "response": "Muito Alto",
                "id": "suicide_final_4"
            }
        }
    }
}
```

Algorithm C.2: Decision Tree Algorithm

```javascript
function calculateDecisionTree(tree, track) {
    if (tree.FinalNode) {
        track.push(tree.id);
        track.push(tree.response);
        return track;
    }
    else {
        response = getResponse(tree.question);
        track.push(tree.question);
        if (response) {
            return calculateRisk(tree.responseYes, track);
        } else {
            return calculateRisk(tree.responseNo, track);
        }
    }
}
```