Predicting the risk associated to pregnancy using data mining

Andreia Brandão, Eliana Pereira, Filipe Portela, Manuel Santos,
António Abelha and José Machado
ALGORITMI Research Centre, Universidade do Minho, Guimarães, Portugal
a60196@alunos.uminho.pt, a58549@alunos.uminho.pt
cfp@dsi.uminho.pt, mfs@dsi.uminho.pt
abelha@di.uminho.pt, jmac@di.uminho.pt

Keywords: Data Mining, Intelligent Decision Support Systems, Voluntary Interruption of Pregnancy, Business Intelligence, Technology Acceptance.

Abstract: Woman willing to terminate pregnancy should in general use a specialized health unit, as it is the case of Maternidade Júlio Dinis in Porto, Portugal. One of the four stages comprising the process is evaluation. The purpose of this article is to evaluate the process of Voluntary Termination of Pregnancy and, consequently, identify the risk associated to the patients. Data Mining (DM) models were induced to predict the risk in a real environment. Three different techniques were considered: Decision Tree (DT), Support Vector Machine (SVM) and Generalized Linear Models (GLM) to perform the classification task. Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology was applied to drive this work. Very promising results were obtained, achieving a sensitivity of approximately 93%.

1 INTRODUCTION

The use of Information and Communication Technologies (ICT) are increasingly, occupying an important place in society. The health sector is no exception, as these among other things, can provide complete and reliable information for healthcare professionals, allowing support to their clinical and administrative decisions and consequently decreasing medical errors associated with these decisions (Pinto, 2009).

An example of a clinical system focused on nursing practice it is the Nursing Practice Support Systems - SSNP a.k.a SAPE. SAPE was created with the goal of giving visibility to the work done by nursing professionals, since these are the users who produce, process and deliver most clinical information (Pinto, 2009).

SAPE is currently used in Centro Hospital of Porto (CHP) and consequently in Centro Materno Infantil do Norte (CMIN) to support nursing practice. In CMIN, the Voluntary Interruption of Pregnancy (VIP) process is a focus of nursing. The patient clinical data records as well as the entire process of VIP are supported by SAPE, making it therefore possible to use this information to extract useful knowledge in the context of Pregnancy Termination (PT).

In this case, it is possible to apply Data Mining (DM) to process the data stored, extracting useful knowledge in order to support clinical decisions based on evidence, as the case of the PT procedure (Bonney, 2013). This process is composed by four steps, where the first step is a medical appointment to verify if the patient is aware of her decision. If she decides to move on, the next two steps are the drug administration and the last step consists of the evaluation process. In this last step of the process, a revision consultation is usually done.

Focusing only on the last step, revision consultation, which is characterized by scheduling a consultation with a physician, four possible situations can occur: VIP was achieved; VIP was not achieved; VIP is incomplete and finally the patient did not appear at the revision consultation.

Having regarded this procedure, the purpose of this article is the use of Data Mining techniques, namely classification models to study the final stage of the VIP process and consequently identify the risk group. This group is characterized by patients who do not appear at the revision consultation and cannot successfully perform the PT procedure.

Besides the introduction, this article includes six sections. The second section is related to the background and related work, which describes the process of VIP and a brief look is taken at the Interoperability System. Subsequently, in section three, the process of Knowledge Discovery in Databases and the method of CRISP-DM, based on the previous is described; the DM techniques used and the statistical metric applied. In the fourth section, each stage of CRISP-DM method is described while the remaining two sections are the discussion and the conclusion.

2 BACKGROUND AND RELATED WORK

2.1 Voluntary Interruption of Pregnancy

In CMIN, the medication methods are used to perform the PT process. This process consists on the administration of specific drugs, in particular *mifepristone* and *misoprostol*.

Furthermore, it is a procedure that involves several steps. The first phase consists in a physician's appointment followed by a period of three days, where the patient needs to consider her decision. Then, the patient receive a medication dosage performed in the CMIN ambulatory. A triage is also executed by the nursing staff in order to verify if the patient is capable to administrate the second medication dosage at home or if she needs to be monitored by the nursing team in CMIN.

After the medication administration phase, a new clinical consultation is performed, where the patient is examined in order to determine whether the procedure was successful. If the opposite occurs, the PT was not achieved or the procedure was incomplete (resulting in the patient admission) (Valente, Cristina, Rosário, & Alcina, 2012).

In this last phase, two situations may occur: the patient is consulted and she is examined by the physician or the patient does not attend to the consultation. In the first case, the patient is evaluated and it is reported if PT was successful. On the other hand, if it was verified ovular remains by ultrasound (incomplete PT), it is necessary to hospitalize the patient. Then if the PT was not achieved, it is necessary to repeat the process.

In the second case, the patient was not evaluated, consequently their condition and the procedure result remains unknown. This situation characterizes who are the risk patients. By the fact that it can originates some problems, in particular, health risks associated to the patient and to the fetus due lacking of medical supervision. Sometimes it constitutes a waste of resources for these patients, since all drugs are reimbursed by the Portuguese State and the entire PT procedure is quite expensive (on average one PT costs 700 euros to the National Health Service).

In maternity care and particular in the VIP module, the use of DM can be considered scarce. So the models studied in this paper can be seen as an innovation in this area and can be used in a Decision Support System for health professionals. To develop this work some Oracle tools (Database and Data Mining) were used.

2.2 Interoperability System

This work, like all the others processes of knowledge extraction in the CHP was possible due to the fact there is in the CHP, the Agency for Integration, Archive and Diffusion of Medical Information (AIDA) is implemented. AIDA is based on the use of intelligent agents, allowing the interoperability among the existing systems. This multi-agent system enables the standardization of clinical systems and overcomes the medical and administrative complexity of the different sources of information from the hospital (Peixoto, Santos, Abelha, & Machado, 2012).

The SAPE resulted from the Nursing Information Systems, as an alternative to the traditional way of information on paper and its design in functional terms. It was developed at the Escola Superior de Enfermagem do São João by Nurse Abel Paiva (Pinto, 2009), (O'Sullivan et al., 2008).

3 KNOWLEDGE DISCOVERY AND DATA MINING

3.1 Knowledge Discovery in Database

The Knowledge Discovery in Database (KDD) is a set of on going activities that enable the extraction of useful knowledge. The main goal of KDD is to discover useful, valid, relevant and new knowledge about a particular activity through algorithms, taking into account the magnitudes of data increasing (Goebel, Siekmann, & Wahlster, 2008).

This process can be resumed in five steps (Cios, 2007):

- Selection: Occurs the selection of the dataset that was needed to run the DM. For this case, data from AIDA and SAPE were used.
- Pre-processing: This step includes cleaning and processing of data, with the aim of making them consistent. For example, the elimination of null values.
- Transformation: At this stage, the data were processed with the ultimate goal of making them uniform. For example, the change in attributes with continuous values for attributes with discrete values;
- Data Mining: According to the type of the desired result, the type of task being performed was defined and the technique to be used was identified. In this case, the classification task was applied and the SVM, GLM and DT techniques were used.
- Interpretation/Evaluation: Consist of the interpretation and evaluation of the patterns obtained (Palaniappan & Awang, 2008).

3.2 Cross Industry Standard Process for Data Mining (CRISP-DM)

The DM methodology that addressed this paper was the Cross Industry Standard Process for Data Mining (CRISP-DM) methodology, because of its characteristics:

- It is an independent tool;
- It is industry independent - The same process can be applied to analyse business, financial data, human resources, manufacturing, services, etc.;
- There is a relationship with the models in the KDD process.

The CRISP-DM breaks the process of Data Mining into six main phases (Catley et al., 2006), (Chattopadhyay et al., 2008), (Frawley et al.,1992): Business Understanding, Data Understanding, Data preparation, Modelling, Evaluation and Implementation.

3.3 Data Mining Techniques

In this work, the following DM techniques were used:

- *Decision Tree (DT)*: DT generates rules automatically, which are conditional statements that reveal the logic used to build the tree.
- *Generalized Linear Models (GLM):* GLM is a statistical technique for linear modelling. GLM is usually implemented for binary classification.
- *Support Vector Machine:* SVM is a powerful algorithm based on linear and nonlinear regression. SVM is usually implemented for binary and multiclass classification.
- *Naïve byes (NB):* is based on applying Bayes' theorem with strong independence assumptions between the features. Although NB also was explored, the obtained results were very weak, consequently there were not presented in this paper.

3.4 Assessment Measures

Through the Confusion Matrix, resulting from the application of DM techniques, four types of results were obtained: a true positive (TP) result that corresponds to the number of positive examples correctly classified; the false positive (FP) result that corresponds to the number of positive examples classified as negative; the true negative (TN) result, that corresponds to the number of negative examples actually classified as negative and, finally, the false negative (FN), that corresponds to the number of negative examples classified as positive.

From this type of values, statistical metrics can be estimated: s*ensitivity,* which is the ability to correctly detect the occurrence of the procedure; *specificity* that is the ability to correctly identify in a model the non-occurrence of a procedure; and *accuracy* that is the total percentage of agreement between the values detected correctly and the actual values (Chapman et al., 2000).

Table 1 presents the expression that characterizes the metrics described.

Table 1: Expressions that define the statistical measures.

Sensitivity	**Specificity**	**Accuracy**
$\frac{TP}{TP+FN}$	$\frac{TN}{TN+FP}$	$\frac{TP+TN}{TP+FP+FN+TN}$

4 DATA MINING PROCESS

CRISP-DM is considered a complex process, but when applied to a certain problem becomes easier to understand, implement and develop. This methodology was used in this case study. Then it is presented all the stages of the process described in section *Cross Industry Standard Process for Data Mining (CRISP-DM).*

4.1 Business Understanding

The process comprises two VIP phases of administering medication that are followed by a medical examination. This examination evaluates whether this process was successful or not. However, there is a group of patients who cannot attend to this appointment and, therefore, are considered a risk group. This is due to not having any information about these patients and the current health state of her or the fetus being unknown.

Thus, a problem was formulated: "*How can be predicted whether a patient belongs to the risk group of patients?".*

This question can be translated into a DM problem: "*What is the probability of a patient belonging to the risk group of patients?*".

A model that predicts whether a woman is considered a risk patient should be constructed based on the data that were previously recorded in the VIP process.

Prior to the development of the model, data associated to the attributes which establish a relationship between the patient and the group of women who did not attend to the evaluation appointment must be selected.

4.2 Data Understanding

This phase includes the extraction of data from SAPE and AIDA using intelligent agents and an analysis of the variables to be used in this DM problem. The extracted data cover the period from 01.01.2012 to 31.12.2012. The amount of data used reaches a number of 1124 records/pregnant. Each record is contains the fields:

- Age: corresponds to the age of the patient who will perform VIP procedure;
- *Number of previous VIP (N_VIP):* this variable sets the number of times the patient underwent the process of VIP previously;
- *Gesta:* corresponds to the number of previous pregnancies of the woman;
- *Para:* corresponds to the number of births that the woman had;
- *Professional Status (PS):* this variable informs if the pregnant woman in question is employed or not.
- *Achieved PT (AA):* this variable informs if the PT procedure was achieved or not.
- *Failed PT (FA):* this variable informs if the PT procedure has failed or not.
- *Incomplete PT (IA):* this variable informs if the PT procedure was incomplete or not.
- *Revision Consultation (RC):* this variable informs if the patient attended to the revision consultation or not;
- *Contraceptive Method (CM):* the variable informs if the pregnant woman had used (1) a contraceptive method or not (0);
- *Weeks of Gestation (WG):* corresponds to the weeks of the gestation of the pregnant woman, when she arrived to CMIN.

A statistical analysis has shown that the data had quality; however they needed to be transformed in order to be incorporated into the DM models.

Figure 1 illustrates the target variable *Revision Consultation* distribution (percentage). As can be seen from this figure, about 10% of patients do not attend to the revision consultation. In the table 2 some statistical measures related to the numerical variables are presented, while in table 3 the percentage of occurrences (number) for each one of the selected variables is shown.

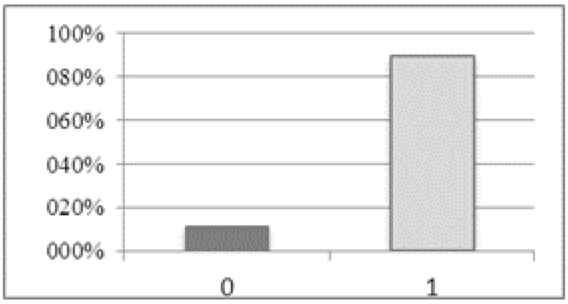

Figure 1. Distribution of values of the target variable *Revision Consultation*, which may take the values attended (1) or not attended (0).

Table 2: Statistics measures of *N_VIP, WG, Age, Gesta* and *Para* variables.

Number	***CM***	***PS***	***N_VIP***	***Para***	***Gesta***
0	38.79	53.11	82.36	48.78	41.58
1	61.21	26.55	14.58	27.90	24.75
2		20.34	2.43	17.82	19.80
3			0.36	4.59	9.18
4			0.27	0.63	2.61
5				0.09	1.08
6				0.09	0.45
8				0.09	0.36
9					0.18

Table 3: Percentage of occurrences of some variables.

	Minimum	Maximum	Average	stDev
N_VIP	0	4	0.22	0.52
WG	3	10.40	7.16	1.46
Age	13	46	27.43	6.95
Gesta	1	9	2.14	1.30
Para	0	8	0.82	0.98

4.3 Data Preparation

At this stage the variables that suited the problem, based on the variables defined in the previous subsection were selected. Thus, the variables *Revision Consultation*, *Gesta*, *Para*, *Professional Status*, *Number of Previous VIPs*, *Age*, *Contraceptive Method* and *Weeks of Gestation* were used in this problem of DM. Subsequently, the selected data were subjected to a pre-processing phase where all the records containing unfilled or noise fields were eliminated. After this processing, there were used only 1119 entries.

Some of the procedures performed to eliminate noise from the data were the replacement of the comma by point to separate the decimals on *Weeks of Gesta*tion variable, the elimination of the text associated to the numerical variables and also the assignment of numerical values to the variable *Professional Status*, where the value 0 is associated with unemployment, the value 1 is associated with employment and value 2 is assigned to the students.

After those transformations of the data, the table of cases with some of the scenarios considered most suitable was built. In a second phase of data preparation, a simple oversampling task was performed in an attempt to obtain better results. Thus, the data was stratified by the target (all rows with 1) and then it was replicated as a package, several times with the goal of balancing the occurrence of target variable (0 and 1). With this operation the dataset maintained their structure.

Thus, the oversampling task consisted of a replication of cases of class "1", i.e. the set of cases "1" has been replicated as many times as necessary until the two elements (“0” and “1”) had a value of approximately cases. This process raised the total number of records up to 1959 (49% for class "0" and 51% for class "1").

4.4 Modelling

In this step, a table of scenarios (table 4) were built, where are represented the 10 scenarios that yielded the best results, arising from different combinations of variables. These scenarios were conceived making use of domain knowledge provided by the clinicians. In each scenario, the target variable *RC* is presented, as well as other variables considered crucial to the creation of forecast models. After defined the table 4, the data were submitted to DM techniques, selected in order to be able to identify the best forecasting model for this DM problem. In this case, the DM techniques used were GLM, SVM and DT.

Table 4: Representation of the variables used in each of the models.

	RC	***Age***	***N_VIP***	***Gesta***	***Para***	***PS***	***CM***	***WG***
Scenario 1	X	X	X	X	X	X	X	X
Scenario 2	X	X	X	X	X	X	X	-
Scenario 3	X	X	X	X	X	X	-	-
Scenario 4	X	X	X	X	X	-	-	-
Scenario 5	X	X	X	-	-	-	-	-
Scenario 6	X	X	X	X	X	X	-	X
Scenario 7	X	X	X	X	X	-	-	X
Scenario 8	X	X	X	X	X	-	X	X
Scenario 9	X	X	X	X	X	-	X	-
Scenario 10	X	X	X	-	-	-	X	X

Each model, resulting from the application of a particular technique for a given DM scenario can be defined by a tuple:

$$M_n = \langle A_f , S_i , TDM_y \rangle$$

The model M_n belongs to the classification approach (A) and it is composed by a scenario (S) {S1 – S10} and a DM technique (TDM) - {GLM, DT, and SVM}. For this DM problem they were induced 60 models were generated, also taking into account the models based on the oversampling dataset. This total amount of models resulted from 10 scenarios x 1 target x 3 TDM x 2 approaches. Table 5 presents the configurations used in the DM models.

Table 5. Configuration of DM Techniques

Technique	**Setting Name**	**Setting Value**	**Setting Type**
DT	Minrec Node	10	Input
	Max Depth	7	Input
	Minpct Split	0.1	Input
	Impurity Metric	Gini	Input
	Minrec Split	20	Input
	Minpct Node	0.05	Input
GLM	Coefficient level	0.95	Default
SVM	Conv tolerance	.001	Input
	Active learning	Enable	Input

4.5 Evaluation

For the evaluation of the obtained results in DM models, an assessment metric has been taken into account, namely sensitivity.

This measure assesses the ability to correctly detect the occurrence of a procedure. It is therefore the most appropriate metric for evaluating the models, since the ultimate goal is to detect the occurrence of patients with a strong probability of belonging to the risk group. At this stage, the top three models were selected for each one of the scenarios and DM techniques.

They are represented in table 6. Using the data obtained from the second phase of data preparation, the techniques of DM used in the first approach were applied again.

For these results, statistical metrics were calculated, including sensitivity, specificity and accuracy for the best models obtained previously (present in the table 6), as shown in the table 7.

Analysing the two tables it is possible to observe that the values of the sensitivity decreased, however the oversampling approach yielded better values for specificity and accuracy. This happens due to a most balanced number of cases in the target variable.

Table 6: Sensitivity, specificity and accuracy values of the top 3 models from the first approach.

SVM		**GLM**		**DT**	
Sensitivity		*Sensitivity*		*Sensitivity*	
Model 4	0.929	Model 4	0.925	Model 4	0.926
Model 7	0.924	Model 5	0.929	Model 5	0.926
Model 10	0.926	Model 9	0.923	Model 9	0.926
Specificity		*Specificity*		*Specificity*	
Model 4	0.100	Model 4	0.093	Model 4	0.090
Model 7	0.093	Model 5	0.092	Model 5	0.090
Model 10	0.101	Model 9	0.092	Model 9	0.090
Accuracy		*Accuracy*		*Accuracy*	
Model 4	0.574	Model 4	0.543	Model 4	0.446
Model 7	0.583	Model 5	0.437	Model 5	0.446
Model 10	0.631	Model 9	0.579	Model 9	0.446

Table 7: Sensitivity, specificity and accuracy values to the top 3 models for each algorithm, using oversampling.

SVM		**GLM**		**DT**	
Sensitivity		*Sensitivity*		*Sensitivity*	
Model 4	0.644	Model 4	0.594	Model 4	0.873
Model 7	0.693	Model 5	0.645	Model 5	0.524
Model 10	0.608	Model 9	0.595	Model 9	0.873
Specificity		*Specificity*		*Specificity*	
Model 4	0.753	Model 4	0.613	Model 4	0.524
Model 7	0.650	Model 5	0.579	Model 5	0.524
Model 10	0.822	Model 9	0.614	Model 9	0.524
Accuracy		*Accuracy*		*Accuracy*	
Model 4	0.680	Model 4	0.602	Model 4	0.555
Model 7	0.670	Model 5	0.605	Model 5	0.555
Model 10	0.657	Model 9	0.603	Model 9	0.555

4.6 Deployment

The extracted knowledge by inducing DM models so as their availability and the way of how the information is presented (through a BI platform), were well accepted by the health professionals. Depending on the requirements, this information may be available through simple or complex reports and can lead to the implementation of a process of repeated DM. In this particular case, the processes of DM are intended to be integrated in the BI platform of CHP.

5 DISCUSSION

After the application of the CRISP-DM methodology, it was possible to verify that the models are quite acceptable, based on the results presented in subsection evaluation. For the induced models the best prediction result obtained based on the sensitivity metric it was approximately 93%. In the table 8, the top 3 models are shown.
As can be seen in table 8, the top three models, taking in consideration the sensitivity values, are the models 4, 5 and 9. Thus, it can be considered that the Decision Tree was one of the best DM techniques applied, considering that the attributes that best characterize the possible patients in the risk group are Age, Number of Previous VIPs, Gesta and Para.

Table 8: The best DM models obtained and the respective technique used.

Models	**DM Technique**	**Sensitivity**
Model 4	Support Vector Machine	0.929
Model 5	Generalized Linear Model	0.929
Model 9	Decision Tree	0.926

Moreover, the results from the second approach, showing a balance in the occurrences of "0" and "1", give more uniform target models, where the ability to correctly detect the patients belonging to a risk group is identical to the ability of detecting patients who do not belong to the same group.

In conclusion, the results obtained for predicting the risk group of patients, based on the presence or absence of the patient in the revision consultation can be considered satisfactory. The oversampling technique reveals not be a good choice because when the models are using sensitive data as it is clinical information it can compromise the data significance.

Thus, the prediction models obtained can be used for supporting decision-making process of the nursing team responsible for the VIP process. Nurses can take into account a priori the characteristics that identify a woman who belongs to the risk group, getting an idea if the patient is a possible "member" of this group, and consequently take preventive measures to these patients.

6 CONCLUSION AND FUTURE WORK

This study demonstrated that it is possible to obtain DM classification models to predict which patient belongs to the risk group in the VIP process. The study was conducted using real data inherent to the VIP process, corresponding to 1 year of activity in CMIN.

Good results were achieved in terms of sensitivity, being approximately 93%, in the model 4 using Support Vector Machines. The most relevant factors to determine the risk group are the patient's age, the number of previous VIPs, the number of previous pregnancies (Gesta) and the number of previous births (Para). It can be concluded that, with the use of classification techniques applied to historical data from pregnant women who underwent in the process of VIP, can be predicted if a patient fits the features of a VIP risk group.

The results are important for a possible redesign of the VIP protocol implemented on CMIN. These models allow improving the service, increasing the number of successes and decrease the costs. At same time can be applied to other institutions with the same problem / protocols.

According to the health professionals these models are very interesting and useful to support the decision when a women wants to make a VIP.

This type of research can promote future developments related with this particular area of maternity care.

ACKNOWLEDGEMENTS

"This work is funded by National Funds through the FCT – Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) within project PEstOE/EEI/UI0752/ 2014 and PEst-OE/EEI/UI0319/2014".

REFERENCES

Abelha, A., Analide, C., Machado, J., Neves, J., & Novais, P. (2007). Ambient Intelligence and Simulation in Health Care Virtual Scenarios, *243*, 461–468.

Bonney, W. (2013). Applicability of Business Intelligence in Electronic Health Record. *Procedia - Social and Behavioral Sciences*, *73*, 257–262. doi:10.1016/j.sbspro.2013.02.050

Catley, C., Frize, M., Walker, C. R., & Petriu, D. C. (2006). Predicting High-Risk Preterm Birth Using Artificial Neural Networks. *IEEE Transactions on Information Technology in Biomedicine*, *10*(3), 540–549. doi:10.1109/TITB.2006.872069

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). The CRISP-DM User Guide. *NCR Systems Engineering Compenhagen*. Brussels: NCR Systems Engineering Copenhagen.

Chattopadhyay, S., Ray, P., Chen, H. S., Lee, M. B., & Chiang, H. C. (2008). Suicidal Risk Evaluation Using a Similarity-Based Classifier. In C. Tang, C. Ling, X. Zhou, N. Cercone, & X. Li (Eds.), *Advanced Data Mining and Applications SE - 7* (Vol. 5139, pp. 51–61). Springer Berlin Heidelberg. doi:10.1007/978-3-540-88192-6_7

Cios, K., Pedrycz, W., Swiniarski, R., & Kurgan, L. (2007). *Data Mining. A knowledge Discovery Approach*. Springer.

Goebel, R., Siekmann, J., & Wahlster, W. (2008). *Advances in Knowledge Discovery and Data Mining*. Springer.

Gonçalves, J., Portela, F., Santos, M. F., Silva, Á., Machado, J., Abelha, A., & Rua, F. (2013). Real-time Predictive Analytics for Sepsis Level and Therapeutic Plans in Intensive Care Medicine. *International Information Institute*.

O'Sullivan, D., Elazmeh, W., Wilk, S., Farion, K., Matwin, S., Michalowski, W., & Sehatkar, M. (2008). Using Secondary Knowledge to Support Decision Tree Classification of Retrospective Clinical Data. In Z. Raś, S. Tsumoto, & D. Zighed (Eds.), *Mining Complex Data SE - 19* (Vol. 4944, pp. 238–251). Springer Berlin Heidelberg. doi:10.1007/978-3-540-68416-9_19

Palaniappan, S., & Awang, R. (2008). Intelligent Heart Disease Prediction System Using Data Mining Techniques. In *Proceedings of the 2008 IEEE/ACS International Conference on Computer Systems and*

Applications (pp. 108–115). Washington, DC, USA: IEEE Computer Society. doi:10.1109/AICCSA.2008.4493524

Peixoto, H., Santos, M., Abelha, A., & Machado, J. (2012). Intelligence in Interoperability with AIDA. In L. Chen, A. Felfernig, J. Liu, & Z. Raś (Eds.), *Lecture Notes in Computer Science, Foundations of Intelligent Systems - 31* (Vol. 7661, pp. 264–273). Springer Berlin Heidelberg. doi:10.1007/978-3-642-34624-8_31

Pinto, L. (2009). *Sistemas de informação e profissionais de enfermagem*. Universidade de Trás-Os-Montes e Alto Douro.

Razavi, A., Gill, H., Åhlfeldt, H., & Shahsavar, N. (2007). Predicting Metastasis in Breast Cancer: Comparing a Decision Tree with Domain Experts. *Journal of Medical Systems*, *31*(4), 263–273. doi:10.1007/s10916-007-9064-1

Valente, C., Cristina, T., Rosário, F., & Alcina, B. (2012). *Acompanhamento de enfermagem na interrupção da gravidez por opção da mulher (I.G.O.)*. Porto.