Application of the Requirements Elicitation Process for the Construction of Intelligent System-Based Predictive Models in the Education Area

Cinthia Vegega(✉), Pablo Pytel, and María Florencia Pollo-Cattaneo

Information System Methodologies Research Group,
Universidad Tecnológica de Buenos Aires, Buenos Aires, Argentina
cinthiavg@yahoo.com.ar, ppytel@gmail.com,
flo.pollo@gmail.com

Abstract. Decision-making is an essential process in the lives of organizations. While each member in an organization makes decisions, this process is particularly important for managerial positions in charge of making decisions on resources allocation. These decisions must be based on predictions about time, effort and/or risks involved in their tasks. Currently, this situation is exacerbated by the complex environment surrounding the organizations, which makes them act beyond their traditional management systems incorporating new mechanisms such as those provided by Artificial Intelligence, leading to the development of an Intelligent Predictive Model. In this context, this work proposes the implementation of a process to assist the Information Systems Engineer in the difficult work of collecting, understanding, identifying and registering the necessary information to implement an Intelligent System-based Predictive Model.

Keywords: Intelligent Systems · Machine Learning · Training data · Education and technology

1 Introduction

What is the most favorable option? What will the future bring us? These are questions we ask ourselves whenever we have to make a decision [1]. For this reason, Humanity has always sought mechanisms to make accurate predictions. Such need not only affects individuals but also organizations. Decision-making is an essential process in the life of organizations. While each member of an organization makes decisions, this process is particularly important for managerial positions. Consequently, managers are known as “decision makers” in their tasks of planning, organizing, directing and controlling [2]. Daily, they have to decide how to allocate valuable resources based on predictions [3] about time, effort and/or risks involved in their tasks. This situation is exacerbated by the highly complex and hardly predictable environment of the 21st century [4], which makes organizations act beyond their traditional management systems and incorporate new mechanisms for the “creation and enhancement of the organization’s knowledge” such as those provided by Artificial Intelligence [5, 6].

H. Florez et al. (Eds.): ICAI 2019, CCIS 1051, pp. 43–58, 2019.
https://doi.org/10.1007/978-3-030-32475-9_4

An example of these new mechanisms can be found in the area of Predictive Models. Despite the fact that Statistical Techniques and Parametric Models have traditionally been used to generate predictions [7], in the last two decades diverse methods associated to Machine Learning [3, 8] have been incorporated. Consequently, it is possible to build models to find a relation between past and future situations using available historical data. In this sense, Artificial Neural Networks [9, 10] and Bayesian Networks [11, 12] can be mentioned as the main Intelligent Systems architectures to be used for this kind of problems [13–15]. These Intelligent System-based Predictive Models possess very useful features, such as generalization capacities, robustness, and self-organization [16].

However, unfortunately, Predictive Models are usually imprecise [3] or, in some cases, they fail, thus often generating incompatible answers [17]. In this regard, the quality of the information required is highly important to make accurate decisions [18]. It is possible to generate more accurate predictions if lack of knowledge on the problem and its context is reduced. Yet, it is almost impossible to have complete, accurate and precise information to make absolutely accurate predictions. There is always a risk related with trusting the available information to assess the situation so a prediction must be associated to a certain degree of probability [19]. Such probability is affected by what is known about the problem and what is not. Consequently, apart from collecting historical data that will be used to build the Predictive Model, it is also necessary to identify the general characteristics of the domain where the prediction is taking place thus being able to detect situations or events of which there are no data but which the model must consider.

In this context, this work proposes the implementation of a process to assist an Information Systems Engineer in the difficult work of collecting, understanding, identifying and registering the necessary information to implement an Intelligent System-based Predictive Model. For that purpose, Sect. 2 presents a brief summary of the proposed process and Sect. 3 presents the results of the implementation in an undergraduate course. Finally, Sect. 4 describes conclusions and future work.

2 Proposed Process

The proposed process aims to assist with the Information Systems Engineers (in their role of Functional Analyst) involved in the implementation of Intelligent Predictive Models, that is, it seeks to support them during the initial phases of the Project considering its particularities. This proposed Project is limited to contemplating the characteristics of two types of Intelligent Systems applied for the implementation of Predictive Models, Multi-layer Perceptron Artificial Neural Networks (ANN) with error Backpropagation training (BPNN) and Bayesian Networks (BN). Consequently, as a result of this proposal, the objectives, success criterion, constraints and assumptions of the Project are determined in order to identify the available information required to train the Intelligent System and to generate an initial specification of it. These results will help the development team start working on the construction, training and validation of the Intelligent Predictive Model to meet the expectations of the organization. The proposed process is structured into the following five phases:

1. **Project Definition Phase:** it aims to define the stakeholders who collaborate in the Project and its scope based on the objectives to be achieved.
2. **Business Process Elicitation Phase:** its objective is to identify and collect the business processes that are relevant for the project, as well as the expert's task in the case of building a model that emulates their prediction capabilities.
3. **Business Process Data Elicitation Phase:** it seeks to identify the data repositories where the information of the different business processes is stored and to collect information about the characteristics of said repositories.
4. **Business Data Conceptualization Phase:** its objective is to identify and evaluate the representativeness of the data available in the business for the construction of the intelligent predictive model.
5. **Intelligent System Initial Specification Phase:** based on the information obtained in the previous phases, the most appropriate type of architecture to implement the predictive model is determined, as well as a proposal of its initial topology.

Figure 1, shown below, presents the proposed process. Each phase of the process defines a set of activities that apply to a case study within the context of an undergraduate course in Sect. 3.

Fig. 1. Phases of the proposed process.

3 Case Study

This section presents the implementation of the phases of the proposed model in a case in a university setting. Firstly, Sect. 3.1 describes the context of the case study, and then describes the application of each phase of the process along with the activities that are carried out in each one of them. Then, the first phase is described in Sect. 3.2, the second phase in Sect. 3.3, the third phase in Sect. 3.4, the fourth phase in Sect. 3.5 and the fifth phase in Sect. 3.6.

3.1 Context of the Case Study

This case study is developed at Facultad Regional Buenos Aires (FRBA), Universidad Tecnológica Nacional (UTN), Argentina. Specifically, it is carried out in the "Systems and Organizations" course [20], of the first year of the "Information Systems Engineering" undergraduate program. The analyzed course is annual and compulsory for students who have passed the admittance course (with approximately 800 enrolled students) and it is one of the integrative courses of the curriculum. In this context, the aim is to implement an Intelligent System to predict the performance of students throughout the course. Because any error in the predictions can lead teachers or students to make wrong decisions, it is of great importance that the system presents consistent results taking into account the normal behavior of the students in the course.

3.2 Application of the First Phase of Project Definition

The following activities are described: "Identify the Objectives of the Project", "Identify the Project stakeholders" and "Identify the Project Scope".

Activity: "Identify the Objectives of the Project"
In this activity, the first conceptual meeting of the project is held with the "Systems and Organizations" Course Chair, which is the main person in charge of the project. The aim is to understand the objective of the project together with the associated expectations. Furthermore, the Chair's faculty members that will take part of the Project's stakeholders are identified, with whom the initial meeting will then be held. Based on the survey carried out, the information obtained is analyzed and the main objectives of the Project are identified, which are documented in the project objectives form, as shown in Fig. 2.

OBJECTIVES OF THE PROJECT		
ID	***Objective Description***	***Priority***
OBJ1	To implement an Intelligent System to predict the performance of the students in the "Systems and Organizations" course taught at UTN FRBA.	High
Observations		
The priority is considered high because it is the only objective that originates the project.		

Fig. 2. Project objectives form.

Activity: "Identify the Project's Stakeholders"
In this activity, the Functional Analyst, based on the information gathered from the organization, identifies the project participants and creates the form shown in Fig. 3.

PROJECT STAKEHOLDERS			
Position	***Org/ Sector***	***Role in the Project***	***Knowledge Areas***
Course Chair	UTN- FRBA / Course	Person in charge	Generalities Course Theory and Practice Pass Requirements
Teacher 1		Stakeholder	
Teacher 2			
TA Monday Course			Course Theory and Practice Data Repositories
TA Tuesday Course			
TA Thursday Course			
TA Friday Course			

where TA means Teaching Assistant

Fig. 3. Project stakeholders form.

Activity: "Identify the Project Scope"
Based on the collected information, the Functional Analyst defines the success criteria of the project as shown in Fig. 4 and determines the problems to be solved in order to

establish what should be included as a result of the project. With this information, the project scope definition form is created, as shown in Fig. 5. This form must be validated by the Course Chair and the business stakeholders. Furthermore, the Functional Analyst also needs to identify the assumptions for the execution of the project. These assumptions include the dependencies on other projects, and all the necessary information that should be available to begin working on the project. This project assumptions form is shown in Fig. 6. Finally, the information restrictions of the project are defined by the project restrictions form, as shown in Fig. 7.

PROJECT SUCESS CRITERIA		
ID	***Criterion Description***	***OBJ-ID***
CE1	To predict the student's performance (including first term exam and make up exams) in the second semester based on data from the first term exam and the first make up exam .	OBJ1

Fig. 4. Project success criteria form.

PROJECT SCOPE DEFINITION		
ID	***Problems to solve***	***OBJ-ID***
P1	To identify students' strengths and weaknesses in order to reinforce what is necessary during the course.	OBJ1
Problems excluded from the Project		
Concept scores and class-to-class evaluations are excluded. Only term exams and make up exams will be considered for the analysis. In addition, the annual planning of the course is not considered for the analysis, being used only for reference of the topics taught.		

Fig. 5. Project scope definition form.

PROJECT ASSUMPTIONS		
ID	***Assumption Description***	***OBJ-ID***
S1	Either by mail or personally, access to information from teaching assistants and teachers will be unrestricted	OBJ1
S2	Data are considered accurate and complete with the same structure since they were provided by the same teacher.	OBJ1

Fig. 6. Project assumptions form.

PROJECT RESTRICTIONS			
ID	***Type***	***Description***	***OBJ-ID***
R1	Data	There are no data on the student's progress class to class.	OBJ1
R2	Data	Assignment scores cannot be used. This is because such data are not considered representative since they are not standardized across all courses.	OBJ1
R3	Data	Students' first names, last names and file numbers cannot be used because they are considered confidential.	OBJ1

Fig. 7. Project Restrictions Form.

3.3 Application of the Second Phase of Business Process Elicitation

The following activities are described: "Identify Business Processes" and "Collect Business Processes". Since the aim is to implement a Predictive Model based on the knowledge of experts available in the organization, the tasks corresponding to the third activity "Collect the Expert's Tasks" are carried out.

Activity: "Identify Business Processes"
From the minutes of the meetings held with project stakeholders, the Project Objectives form (Fig. 2), the Project Success Criteria form (Fig. 3) and the Project Scope Definition form (Fig. 4), the Functional Analyst defines the most significant business activities for the project and makes a use case diagram that is included in the business process diagram form (Fig. 8).

BUSINESS PROCESS DIAGRAM				
Business Process List				
ID	***Name***	***Description***	***Person in Charge***	***OBJ-ID***
P0001	Evaluate Students' Exams	The aim of this process is to correct the students' exams and load the data into repositories intended for this purpose.	Teaching Assistant	OBJ1
Main Diagram				

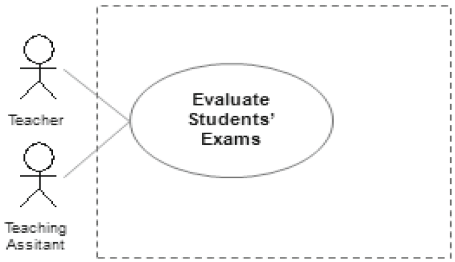

Fig. 8. Business process diagram form.

Activity: "Collect Business Processes"

Taking into account the collected information associated with the identified business process, the Functional Analyst holds a new meeting with the Course Chair and the course stakeholders. In this way, information is collected to record how this process works and how it is related to data repositories. From the information gathered, the information is registered by documenting it in the business process form, as shown in Fig. 9.

BUSINESS PROCESSES	
P0001 – Evaluate Students' Exams	
Description	The aim of this process is to correct the students' exams and load the data into repositories intended for this purpose.
Actors	Teacher; Teaching Assistant (TA)
Pre-Conditions	Students sat for the exams to be evaluated and they were already distributed among teachers and teaching assistants for their correction
Post-Conditions	The exams were corrected the scores were registered in the course spread-sheet.
Normal Flow	
1	The TA corrects the practice part of the term exams.
2	The teacher corrects the theory part of the term exams.
3	The TA holds a meeting with the teacher to deal with doubts during the correction.
4	The teacher determines the term exam final score.
5	The TA registers the scores in the course spreadsheet.
Alternative Flow	
1	In the event that an exam has been inaccurately corrected after the student has analyzed it, the teacher and the teaching assistant will meet again in order to correct it again and define the final score.
2	Once the score is determined, the teaching assistant registers it in course spreadsheet.

Fig. 9. Business process form.

Activity: "Analyze the Expert's Tasks"

From previous meetings, it has been detected that the task of determining whether the student passes or not taking into account their characteristics and the result of their exams depends on different types of knowledge that is internalized in the minds of teachers. This means that there are no fixed rules or standard procedures to carry out such task so it is not possible to identify a Business Process. Therefore, it is an expert task and it is decided to analyze it in order to obtain the knowledge applied by the Course Chair. To do this, the Protocol Analysis is selected as the knowledge elicitation technique, thus performing the steps corresponding to the technique according to [21]. Once the steps have been carried out based on the protocol, the Functional Analyst will register the knowledge obtained in the expert's task form, as shown in Fig. 10a and b.

(a)

EXPERT'S TASKS	
TE001 – Evaluate Course Academic Status of the Student	
General Description	The objective is to determine whether the student will or will not pass the course considering their exam results.
Factual Knowledge	

Table of Concepts – Characteristics – Values

Concepts	*Characteristics*	*Values*
Student	*FileNumber_Student*	Alphanumeric
	FullLastName_Student	Alphanumeric
	FullName_Student	Alphanumeric
Evaluation	*Type_Evaluation*	Alphanumeric
	FinalScore_Theory	Alphanumeric
	FinalScore_Práctice	Alphanumeric

Glossary of Terms

Term	Description
Evaluation	It indicates the exam the student is sitting for.
FileNumber_Student	Student's File Number (assigned by the School).
FinalScore_Práctice	Final Score obtained in the practice part of the exams.
FinalScore_Theory	Final Score obtained in the theory part of the exams.
FullLastName_Student	Student's Full Last Name.
FullName_Student	Student's Full Name.
Student	It indicates the student who takes the course, sits for the exams and submits the assignments.
Type_Evaluation	Type of exam, which may be first term exam, second term exam or make-up exams.

Tactical Knowledge

The rules resulting from the Protocol Analysis performed with the Course Chair are as follows:

- If the student passes the theory of the first term exam or the theory of the first make-up of the first term exam or the theory of the second make-up of the first term exam, then the student passes the theory of the first semester.
- If the student passes the practice of the first term exam or the practice of the first make-up of the first term exam or the practice of the second make-up of the first term exam, then the student passes the practice of the first semester.
- If the student passes the theory of the second term exam or the theory of the first make-up of the second term exam or the theory of the second make-up of the second term exam, then the student passes the theory of the second semester.
- If the student passes the practice of the second term exam or the practice of the first make-up of the second term exam or the practice of the second make-up of the second term exam, then the student passes the practice of the second semester.
- If the student passes the theory of the first semester and the practice of the first semester and the theory of the second semester and the practice of the second semester, then the student passes the course.
- If the student does not pass the theory of the first term exam, then it is likely that they will not pass the theory of the second term exam.
  If the student does not pass the practice of the first term exam, then it is likely that they will not pass the practice of the second term exam.

Fig. 10 **a.** Expert's task form (factual and tactical knowledge) **b.** Expert's task form (strategic knowledge and metaknowledge).

(b)

EXPERT'S TASKS
TE001 – Evaluate Course Academic Status of the Student
Strategic Knowledge
A three-level decomposition tree is drawn in https://bit.ly/2wdk9JF
Metaknowledge
A knowledge map is drawn in https://bit.ly/2HRBz3W

Fig. 10 (*continued*)

3.4 Application of the Third Phase of Business Process Data Elicitation

The following activities are described: "Identify Data Repositories" and "Collect Business Data".

Activity: "Identify Data Repositories"

The Functional Analyst analyzes the information gathered from the interviews conducted with the business stakeholders and from the "Expert's Tasks" form (Fig. 10). As a result, he detects that the main data to be used in the project are in Excel format spreadsheets named *planning_ < course day >* which describe the students' behavior (in relation to their exam, make-up and assignment scores) during the year. These spreadsheets were provided by the Course Chair. After defining such data repositories, the Functional Analyst registers this information and prepares the data repository form shown in Fig. 11.

DATA REPOSITORIES					
ID	***Name***	***Type***	***Description***	***Business Process /Task***	***Person in charge***
P_M	planning_ Monday.xls	Excel spreadsheet	Evaluation results of Monday course students.	P0001; TE001	TA Monday course
P_Tu	planning_ Tuesday.xls	Excel spreadsheet	Evaluation results of Tuesday course students.	P0001; TE001	TA Tuesday course
P_Th	planning_ Thursday.xls	Excel spreadsheet	Evaluation results of Thursday course students.	P0001; TE001	TA Thursday course
P_Fr1	planning_ Friday 1M.xls	Excel spreadsheet	Evaluation results of Friday course 1° module students.	P0001; TE001	TA Friday course
P_Fr2	planning_ Friday 2M.xls	Excel spreadsheet	Evaluation results of Friday course 2° module students.	P0001; TE001	TA Friday course

Fig. 11. Data repository form.

Activity: "Collect Business Data"
The Functional Analyst prepares the data structure form, shown in Fig. 12. Since all the Excel spreadsheets provided have the same format, a generic data structure is obtained, which is valid for all the data repositories.

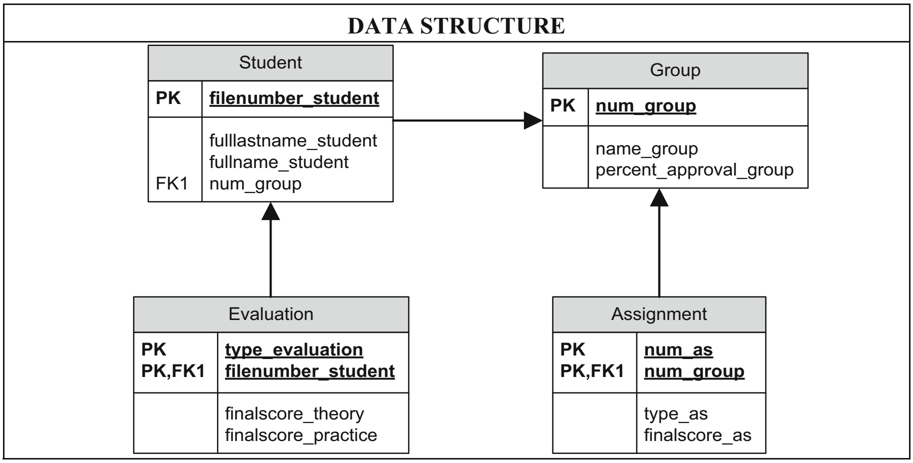

Fig. 12. Data structure form.

3.5 Application of the Fourth Phase of Business Data Conceptualization

The purpose of this phase is to determine whether the data that will be used in the Predictive Model are representative. To do this, in this case study, the two cycles of this phase are conducted, first evaluating the more general initial version of the business data and then a more detailed extended version.

Evaluation of the Initial Version of the Business Data
The evaluation of the initial version of the data is made, completing the tasks corresponding to the activity named "Identify Data to Build the Predictive Model" and then those corresponding to the second activity, i.e. "Validate Data Representativeness".

Activity: "Identify Data to Build the Predictive Model"
Based on the "Data Repository" form (Fig. 11), the "Data Structure" form (Fig. 12) and the information collected in the interviews conducted to the business stakeholders, the Functional Analyst documents the obtained data in the "Available Data" form shown in Fig. 13. In this case, all the spreadsheets identified as data repositories were integrated into a single spreadsheet which includes all the data.

Activity: "Validate Data Representativeness"
Based on the available data obtained in the previous activity, the Functional Analyst analyzes whether such data are representative of the business in order to build the Predictive Model. Due to space constraints, all the tasks performed on the data set are specified in [22]. The conclusions drawn show that the initial version of the data (more

AVAILABLE DATA

Data Version	Initial Version of the Data (1.0)

The data repositories named *planning_<course day>.xls* have three tabs: “Annual Planning”, “Assignments” and “Term Exams”. For this project, only the “Term Exams” tab is used, since the data structure includes neither the data related to the assignments nor the data related to the groups of students doing such assignments. This is because such data are not considered representative since they are not standardized across all courses. In addition, students’ first names, last names and file numbers cannot be used because they are considered confidential. Both situations are accounted for in the Project Restrictions form (figure 6). After integrating the 5 spreadsheets, identified in figure 10, a single spreadsheet is obtained, from which a sample of 75 records is taken. The integrated data are specified, which includes 14 attributes corresponding to 75 records containing the students’ performance in the theory and practice parts of the course exams (term exams and make-ups) as well as the course day and the resulting course academic status.
In this case, two attributes are described as an example.

Attribute:	*COURSE ACADEMIC STATUS*
Description:	Course result, that is, whether the student has passed the course or not or they directly do not appear in the Final Course Records (TPA) due to absenteeism.
Type of Data:	Alphanumeric
Value Range:	- APRUEBA - NO_APRUEBA - NO_TPA

Attribute:	*C1_P_P; C1_1R_P; C1_2R_P*
Description:	Final score of the practice part of the exams (it includes all the practice items of the first semester, which are Process Charts and Organizational Charts), where the prefix means the following: • C1_P: First Term Exam • C1_1R: First Make-up of the First Term Exam • C1_2R: Second Make-up of the First Term Exam
Type of Data:	Alphanumeric.
Value Range:	- AP_dist: Passed with Honors - AP: Satisfactory Pass - AP_err: Passed with some errors - AP_lim: Borderline Pass - NO_AP_lim: Borderline Fail - NO_AP: Failed - NC: Topic Not Answered - AUS: Student Absent in the Evaluation - NA: Topic that the student does not need to answer (this value is used for make-up exams only).
Synonym:	First Semester Practice Score

Fig. 13. Available data form for the initial version of the data.

general and limited) contains biases which generate differences with the relationships preconceived by the teacher of the courses. For this reason, a new version of the data is generated, including more examples and more detail in the attributes used.

Evaluation of the Extended Version of the Business Data
The tasks corresponding to the activity named "Identify Data to Build the Predictive Model" and those corresponding to the second activity, "Validate Data Representativeness" are performed in order to evaluate the extended version of the data.

Activity: "Identify Data to Build the Predictive Model"
In view of the problems detected in the initial version of the data, it is decided to extend them both in number of rows and in detail of the attributes, which are also documented in a new "Available Data" form, shown in Fig. 14.

AVAILABLE DATA	
Data Version	Extended Version of the Data (2.0)
In order to extend the data, a detailed account of the results for each topic of each exam is made (in this way, for instance, in the Practice part of the first term exam, the results for Organizational Charts and Process Charts are indicated). In addition, 51 students' records are added so that the data include a total of 126 rows with 23 attributes. Below, an attribute is described in detail as an example.	

Attribute:	*C1_P_ORG; C1_1R_ORG; C1_2R_ORG*
Description:	Final exam score for the organizational charts topic which is evaluated in the first semester, where the prefix means the following: • C1_P: First Term Exam • C1_1R: First Make-up of the First Term Exam • C1_2R: Second Make-up of the First Term Exam
Type of Data:	Alphanumeric.
Value Range:	- AP_dist: Passed with Honors - AP: Satisfactory Pass - AP_err: Passed with some errors - AP_lim: Borderline Pass - NO_AP_lim: Borderline Fail - NO_AP: Failed - NC: Topic Not Answered - AUS: Student Absent in the Evaluation - NA: Topic that the student does not need to answer (this value is used for make-up exams only).
Synonym:	Score obtained in Organizational Charts in the First Semester

Fig. 14. Available data form for the extended version of the data.

Activity: "Validate Data Representativeness"
Based on the available data obtained in the previous activity, the Functional Analyst once again analyzes whether such data are representative of the business in order to build a Predictive Model. Therefore, all the tasks performed on the data set are

described in detail. All the tasks performed on the data set are specified in [22]. Based on the conclusions obtained, the extended version is representative of the students' behavior and therefore it is the version that will be used to build the Predictive Model.

3.6 Application of the Fifth Phase of Initial Specification of the Intelligent System

The following activities are described: "Select the Type of Intelligent System" and "Define Initial Topology of the Intelligent System". In addition, the complete process is specified in [23].

Activity: "Select Type of Intelligent System"
The Functional Analyst answers the questions associated to each characteristic using the meeting minutes written in previous phases, formalizing them in Table 1 as shown below. The possible linguistic values that can be used for each characteristic are "Nothing", "Little", "Regular", "Much" and "All". Once the linguistic values corresponding to each characteristic defined in Table 1 are assigned, the Functional Analyst obtains the values corresponding to each architecture and selects the best architecture for the project. The operations made are presented in the spreadsheet available in [24]. As shown in Fig. 15, the architecture selected in this case is Bayesian Networks. This selection is registered in the predictive model architecture form, shown in Fig. 16.

Table 1. Characteristics evaluated in order to define the most appropriate architecture.

Category	ID	Question associated to the characteristic	Value
Available data	D1	How much confidence is there as to the representativeness of the data?	Much
	D2	To what extent may the data be considered complex and with a nonlinear relationship between their attributes?	Little
	D3	How many examples do the data include?	Regular
	D4	What percentage of data is there with continuous numeric values (in relation to non-numeric values or numeric discrete values)?	Regular
Expected results	R1	To what extent is prediction accuracy considered critical?	Much
	R2	To what extent is it desirable to know and compare the predictions for different possibilities and scenarios?	Much
	R3	To what extent is it important to be able to explain how the results generated were obtained?	Much
Problem domain	P1	How stable is the problem to be solved?	All
	P2	To what extent are the domain experts available to participate in the project?	All
	P3	To what extent it is desirable to be able to manually adjust the network based on the knowledge about the data?	All

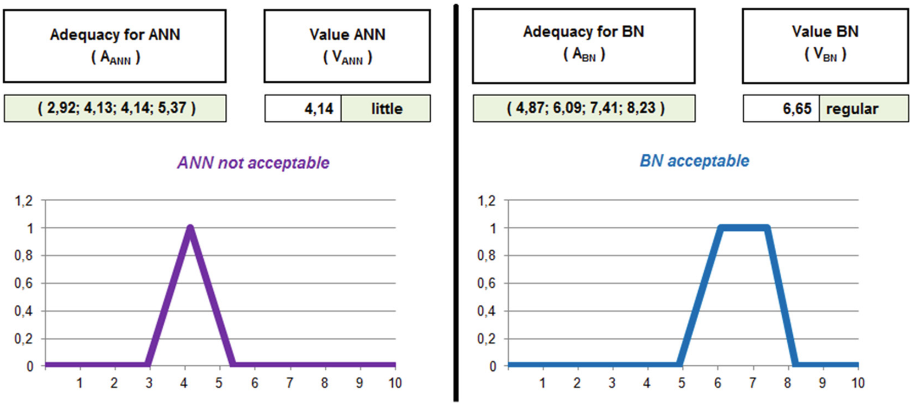

Fig. 15. Appropriateness values for each architecture.

PREDICTIVE MODEL ARCHITECTURE
The most appropriate architecture for this project is Bayesian Networks, while the use of Artificial Neural Networks is discarded for the problem to be solved. The available data include a sufficient number of examples to perform the training and validation of the network, allowing the definition of value ranges for each of them. In addition, there are domain experts available who can contribute their knowledge and participate in the project. Therefore, comparisons between the predictions for different scenarios will be possible as well as knowing how the results were obtained, thus allowing for adjustments in the network that is defined.

Fig. 16. Predictive model architecture form.

Activity: "Define Initial Topology of the Intelligent System"
Once the type of architecture to be used is selected, the Functional Analyst defines the initial characteristics of the topology and documents them in the initial topology of the predictive model architecture form.

For building the model proposed, the Analyst uses the 'Graphical Network Interface' or GeNIe software tool [25], together with the extended version of the data obtained in the previous phase. A simple structure is chosen to be applied, where each node connects with the objective attribute (COURSE ACADEMIC STATUS) and, in the case of the topics, the result of the term exam is associated to the first make-up and the latter with the second make-up. The reason for this linkage is that, as explained by the domain expert, it is thought that the result that a student would obtain in the theory part of the second make-up may be influenced by the results of the first make-up and of the term exam. With such structure, the available data (corresponding to the Extended Version of the Data) are imported so that the tool can determine the a priori probabilities and the conditional probabilities corresponding to each node. As a result of this operation, the probability distribution is obtained, as shown in https://bit.ly/2W1X1gd (the names are displayed in the original language).

Although the implementation of the final Bayesian Network falls outside the scope of this process, in order to confirm whether this initial topology was successfully trained, the validation thereof was performed using the same data with a functionality provided by GeNIe. Despite the fact that in a real project it would not make sense to validate an Intelligent System using the same data as those used to train it (since the accuracy thus obtained is not reliable), in this case we only seek to confirm that the probabilities given by the network can be considered representative of the data used. As a result, the general accuracy of this network is 94%, with 100% of accuracy to predict students that approve the course, 87.5% for students that do not approve and 96,4% for students that do not finish the course. Therefore, it is possible to affirm that this initial topology is useful to be used as a basic prototype of the Intelligent Model to predict the students' behavior in the course.

4 Conclusions

In this work, a proposed process has been applied to predict the performance of students throughout a university course. In the first phase, the objective and requirements of the Predictive Model has been defined. In the second phase, the business process characteristics have been identified. During the third phase, the available data sources have been detected, which then have been evaluated in the fourth phase to retrieve a data set sufficiently representative of the behavior of the students. Finally, in the fifth phase the most appropriate technology to build the Predictive Model has been established to satisfy the project requirements.

References

1. De Kohan, N.C.: Los sesgos cognitivos en la toma de decisiones. Int. J. Psychol. Res. **1**(1), 68–73 (2008). ISSN 2011-7922
2. Robbins, S., Coulter, M.: Administración, Décima Edición. Prentice Hall, Upper Saddle River (2010). ISBN: 978-607-442-388-4
3. Mair, C., et al.: An investigation of machine learning based prediction systems. J. Syst. Softw. **53**(1), 23–29 (2000)
4. García, F.J.M., Martínez, M.A.P., García, J.S.: Gestión Estratégica del Conocimiento. Asociación Universitaria Iberoamericana de Postgrado (2003)
5. Nilsson, N.J.: Principles of Artificial Intelligence. Morgan Kaufmann, Burlington (2014)
6. Russell, S.J., Norvig, P., Davis, E., Russell, S.J., Russell, S.J.: Artificial Intelligence: A Modern Approach, vol. 2. Prentice Hall, Englewood Cliffs (2010)
7. Shepperd, M., Kadoda, G.: Comparing software prediction techniques using simulation. IEEE Trans. Softw. Eng. **27**(11), 1014–1022 (2001)
8. Bontempi, G., Ben Taieb, S., Le Borgne, Y.-A.: Machine learning strategies for time series forecasting. In: Aufaure, M.-A., Zimányi, E. (eds.) eBISS 2012. LNBIP, vol. 138, pp. 62–77. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-36318-4_3
9. Wang, S.-C.: Artificial neural network. In: Wang, S.-C. (ed.) Interdisciplinary Computing in Java Programming. SECS, vol. 743, pp. 81–100. Springer, Boston (2003). https://doi.org/10.1007/978-1-4615-0377-4_5

10. Wang, L., Fu, K.: Artificial neural networks. In: Wiley Encyclopedia of Computer Science and Engineering, pp. 181–188 (2009)
11. Barber, D.: Bayesian Reasoning and Machine Learning. The MIT Press, Cambridge (2012)
12. Premchaiswadi, W. (ed.): Bayesian Networks. In-Tech. (2012)
13. Chatfield, C.: The Analysis of Time Series: An Introduction. CRC Press, Boca Raton (2016)
14. De Gooijer, J.G., Hyndman, R.J.: 25 years of time series forecasting. Int. J. Forecast. **22**(3), 443–473 (2006)
15. Zhang, G., Hu, M.Y.: Neural network forecasting of the British pound/US dollar exchange rate. Omega Int. J. Manag. Sci. **26**(4), 495–506 (1998)
16. Cohen, P.R., Feigenbaum, E.A.: The Handbook of Artificial Intelligence, vol. 3. Butterworth-Heinemann, Oxford (2014)
17. Acquatela, H.: La predicción del futuro: desde el oráculo de Delfos hasta la medicina actual. Gac. méd. Caracas **114**(2), 150–156 (2006). ISSN 0367-4762
18. Rodríguez, M., Márquez Alegría, M.: Manejo de problemas y toma de decisiones, vol. 8. Editorial El Manual Moderno (2015). SBN: 9789684264670
19. Arsham, H.: Tools for decision analysis: analysis of risky decisions (2006)
20. Universidad Tecnológica Nacional – Facultad Regional Buenos Aires. Programa de la carrera Ingeniería en Sistemas de Información (Plan 2008). DISI (2008). https://tinyurl.com/y2kb36xv
21. Gómez, A., Juristo, N., Montes, C., Pazos, J.: Ingeniería del Conocimiento. Centro de Estudios Ramón Areces. S.A. Madrid (1997)
22. Vegega, C., Pytel, P., Straccia, L., Pollo-Cattaneo, M.F.: Evaluation of the bias of student performance data with assistance of expert teacher. In: Florez, H., Diaz, C., Chavarriaga, J. (eds.) ICAI 2018. CCIS, vol. 942, pp. 16–31. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01535-0_2. ISSN: 1865-0937
23. Vegega, C., Pytel, P., Straccia, L., Pollo-Cattaneo, M.F.: Proceso de Selección de Arquitectura a fin de Implementar un Modelo Predictivo Inteligente. Memorias de 6to. Congreso Nacional de Ingeniería Informática y Sistemas de Información (CONAIISI 2018). Workshop de Aplicaciones Informáticas y de Sistemas de Información (2018). https://tinyurl.com/y6nzhjnv. ISSN 2347-0372
24. Vegega, C., Pytel, P., Pollo-Cattaneo M.F.: Método Evaluador de Arquitectura - Desempeño de Alumnos de Sistemas y Organizaciones (2017). https://bit.ly/2YwE5UM
25. BayesFusion: BayesFusion - Home Page for users of GeNIe. BayesFusion LLC (2015). https://www.bayesfusion.com/