

Elicitación de Requerimientos para la Construcción de Modelos Predictivos basados en Sistemas Inteligentes dentro del Ámbito Educativo

Cinthia Vegega^{1,2}, Pablo Pytel^{1,2} & María Florencia Pollo-Cattaneo^{1,2}

¹ Grupo de Estudio en Metodologías de Ingeniería en Software (GEMIS). Universidad Tecnológica Nacional. Facultad Regional Buenos Aires. Argentina.

² Programa de Maestría en Ingeniería en Sistemas de Información. Facultad Regional Buenos Aires. Universidad Tecnológica Nacional. Argentina

cinthiavg@yahoo.com.ar; pppytel@gmail.com; flo.pollo@gmail.com

Resumen. La toma de decisiones es un proceso esencial en la vida de las organizaciones siendo particularmente importante para los roles gerenciales que deben tomar decisiones sobre cómo asignar sus recursos. Estas decisiones deben basarse en predicciones sobre el tiempo, el esfuerzo y/o los riesgos de llevar a cabo sus actividades. Actualmente esta situación se encuentra acentuada por el entorno complejo que rodea a las organizaciones lo cual provoca que deban actuar más allá de los sistemas tradicionales de gestión incorporando nuevos mecanismos, como los provistos por la Inteligencia Artificial, por lo que se estaría desarrollando un Modelo Predictivo Inteligente. En este contexto, el presente trabajo tiene como objetivo proponer la aplicación de un proceso que asista al Ingeniero en Sistemas de Información en la difícil tarea de recolectar, entender, identificar y documentar la información necesaria para implementar un Modelo Predictivo basado en Sistemas Inteligentes.

Palabras Claves: Sistema Inteligente. Modelo Predictivo. Educación y Tecnología. Toma de Decisiones. Organización.

1. Introducción

La toma de decisiones implica cuestionarse la validez de las alternativas, sus ponderaciones, los criterios de selección a utilizar y su impacto a futuro [1]. Por esta razón, contar con mecanismos que permitan realizar predicciones acertadas es algo que la Humanidad siempre ha querido. Dicha necesidad no es propia únicamente de los individuos y también impacta en las organizaciones. La toma de decisiones es un proceso esencial en la vida de las organizaciones. Aunque cada miembro dentro de una organización toma decisiones, este proceso es particularmente importante para los roles gerenciales. Es por esto que a los gerentes se los conoce como “tomadores de decisiones” en sus actividades de planear, organizar, dirigir y controlar [2]. Diariamente deben decidir cómo asignar sus valiosos recursos basados en predicciones [3] sobre el tiempo, el esfuerzo y/o los riesgos que implica llevar a cabo sus actividades. Esta situación se encuentra acentuada debido al entorno altamente complejo y de difícil predicción del siglo XXI [4], lo cual genera que las organizaciones deban actuar más allá de los sistemas tradicionales de gestión e incorporar nuevos mecanismos para la “creación y potenciación del conocimiento organizativo” tales como los provistos por la Inteligencia Artificial [5; 6].

Un caso de estos nuevos mecanismos se puede encontrar en el ámbito de los Mo-

delos Predictivos. A pesar de que tradicionalmente se han aplicado Técnicas Estadísticas y Modelos Paramétricos para generar predicciones [7], en las dos últimas décadas se han incorporado diversos métodos asociados al Aprendizaje Automático [8; 3]. De esta manera, es posible construir modelos que permitan encontrar la relación entre situaciones pasadas y futuras usando los datos históricos disponibles. En este sentido, se pueden destacar las Redes Neuronales Artificiales [9; 10] y las Redes Bayesianas [11; 12] como las principales arquitecturas de Sistemas Inteligentes a ser utilizadas para este tipo de problema [13; 14; 15]. Estos Modelos Predictivos basados en Sistemas Inteligentes presentan características muy útiles, tales como son la capacidad de generalización, robustez y auto-organización [16].

No obstante, desafortunadamente los Modelos Predictivos suelen ser imprecisos [3] o, en algunos casos, se equivocan generando a menudo respuestas que son incomprensibles [17]. En este sentido, la calidad de la información requerida es sumamente importante para poder tomar decisiones bajo certeza [18]. Al reducir la ignorancia sobre el problema y su contexto es posible generar mejores predicciones. Sin embargo, contar con información completa, cierta y precisa para predecir con total certeza es casi imposible. Siempre existe un riesgo asociado en confiar en la información disponible para evaluar la situación por lo que una predicción deberá tener asociado cierto grado de probabilidad [19]. Dichas probabilidades se encuentran afectadas por lo que se conoce del problema, como por lo que no. Por consiguiente, además de recolectar los datos históricos que se aplicarán para construir el Modelo Predictivo, también es imprescindible identificar las características generales del dominio en donde se lleva a cabo la predicción, y así poder detectar situaciones o eventos de los que no se posee datos pero que el modelo deberá considerar.

En este contexto, el presente trabajo aplica un proceso propuesto para asistir al Ingeniero en Sistemas de Información en la difícil tarea de recolectar, entender, identificar y documentar la información necesaria para implementar un Modelo Predictivo basado en Sistemas Inteligentes. Para ello, en la sección 2 se presenta un breve resumen del proceso propuesto y en la sección 3 se indican los resultados de su aplicación en un caso dentro del contexto de una asignatura de grado. Finalmente, en la sección 4 se describen las conclusiones.

2. Proceso Propuesto de Elicitación de Requerimientos

El proceso propuesto busca ser una guía para los Ingenieros en Sistemas de Información involucrados en la implementación de Modelos Predictivos Inteligentes, que brinda soporte durante las etapas iniciales del proyecto teniendo presente las particularidades del mismo. Este proceso se limita a contemplar las características de dos tipos de Sistemas Inteligentes aplicables para la implementación de Modelos Predictivos: las Redes Neuronales Artificiales (RNA) de tipo Multi-Perceptrón con aprendizaje por retropropagación del error hacia atrás (también conocidas como RNA Multi-Perceptrón Backpropagation) y, las Redes Bayesianas (RB). Por lo tanto, como resultado de esta propuesta se determinan los objetivos, criterio de éxito, restricciones y suposiciones del proyecto lo que permite identificar los datos disponibles que serán necesarios utilizar para el entrenamiento del Sistema Inteligente y generar una especificación inicial del mismo. A partir de estos resultados, el equipo de desarrollo podrá comenzar a trabajar en la construcción, entrenamiento y validación del Modelo Predi-

cativo Inteligente que cumpla las expectativas de la organización. El proceso propuesto se estructura en las siguientes cinco fases:

- 1. Fase de Definición del Proyecto:** tiene como objetivo, definir los interesados que colaboran en el proyecto y el alcance del mismo a partir de los objetivos que se desean alcanzar.
- 2. Fase de Educción de Procesos de Negocio:** su objetivo es identificar y relevar los procesos de negocio que son significativos para el proyecto, así como las tareas del experto en el caso de construir un modelo que emule sus capacidades de predicción.
- 3. Fase de Educción de Datos de Procesos de Negocio:** busca identificar los repositorios de datos donde se almacena la información de los diferentes procesos de negocio y relevar las características de dichos repositorios.
- 4. Fase de Conceptualización de los Datos del Negocio:** su objetivo es identificar y evaluar la representatividad de los datos disponibles en el negocio para la construcción del modelo predictivo inteligente.
- 5. Fase de Especificación Inicial del Sistema Inteligente:** a partir de la información obtenida en las fases anteriores, se determina el tipo de arquitectura más apropiada para implementar el modelo predictivo, así como, una propuesta de la topología inicial del mismo.

Cada fase del proceso tiene definido un conjunto de actividades que se aplican a un caso de estudio dentro del contexto de una asignatura de grado en la sección 3 (la explicación detallada de cada actividad del proceso se encuentra disponible en [20]).

3. Aplicación del Proceso Propuesto en un Caso de Estudio

En esta sección se presenta la aplicación de las fases del proceso propuesto en un caso que pertenece al dominio de una universidad. En primera instancia, en la sección 3.1 se describe el contexto del caso de estudio, para luego describir la aplicación de cada fase del proceso junto con las actividades que se llevan a cabo en cada una de ellas. Así, la primera fase se describe en la sección 3.2, la segunda en la sección 3.3, la tercera fase en 3.4, la cuarta fase en 3.5 y la quinta fase en la sección 3.6.

3.1. Contexto del Caso de Estudio

Este caso de estudio se desarrolla dentro del contexto de la Facultad Regional Buenos Aires (FRBA) perteneciente a la Universidad Tecnológica Nacional (UTN) de la República Argentina. Específicamente, se realiza en la asignatura “Sistemas y Organizaciones” [21], perteneciente al primer año de la carrera de “Ingeniería en Sistemas de Información”. La asignatura analizada es anual y de cursado obligatorio para los alumnos que hayan aprobado el curso de ingreso a la carrera (con aproximadamente 800 inscriptos) y se corresponde al tronco integrador del Plan de Estudios. En este contexto, se busca implementar un Sistema Inteligente que permita predecir el desempeño de los alumnos a lo largo de la cursada de dicha materia. Debido a que cualquier error en las predicciones puede llevar a que los docentes, o alumnos, tomen decisiones equivocadas, se considera de gran importancia que el sistema presente resultados consistentes teniendo en cuenta el comportamiento normal de los alumnos de la asignatura.

3.2. Aplicación de la Primera Fase de Definición del Proyecto

Las siguientes actividades son descriptas: “Identificar los Objetivos del Proyecto”

(sección 3.2.1), “Identificar los Interesados del Proyecto” (sección 3.2.2), y “Identificar el Alcance del Proyecto” (sección 3.2.3).

3.2.1. Actividad “Identificar los Objetivos del Proyecto”

En esta actividad se lleva a cabo la primera reunión conceptual del proyecto con la jefa de cátedra de la asignatura “Sistemas y Organizaciones”, que es la principal patrocinadora del proyecto. De esta manera, se busca entender en forma general el objetivo del proyecto junto con las expectativas en relación al mismo. Asimismo, se identifican los profesores de la cátedra y los ayudantes que conformarán el conjunto de interesados del proyecto, con los cuales luego se realiza la reunión de inicio. A partir del relevamiento realizado se analiza la información obtenida e identifican los principales objetivos del proyecto, que se documentan en el formulario de objetivos del proyecto, tal como se muestra en la figura 1.

3.2.2. Actividad “Identificar los Interesados del Proyecto”

En esta actividad, el Analista Funcional, a partir de la información relevada de la organización, identifica a los participantes del proyecto y confecciona el formulario que se muestra en la figura 2.

3.2.3. Actividad “Identificar el Alcance del Proyecto”

A partir de la información recolectada, el Analista Funcional define los criterios de éxito del proyecto tal como se muestra en la figura 3 y releva los problemas a resolver a fin de establecer qué es lo que se debe incluir como resultado del proyecto. Con esta información, se construye el formulario de definición de alcance del proyecto, tal como se muestra en la figura 4. Dicho formulario debe ser validado por el patrocinador e interesados del negocio. Asimismo, identifica cuáles son los supuestos que el proyecto considera como base para la ejecución del mismo, y cuáles son las dependencias de otros proyectos o información que se consideran necesarios que estén disponibles para poder comenzar a trabajar en el proyecto. Esto se incluye en el formulario de suposiciones del proyecto, tal como se muestra en la figura 5. Por último, se definen las restricciones de información del proyecto y se confecciona el formulario de restricciones del proyecto, tal como se muestra en la figura 6.

OBJETIVOS DEL PROYECTO		
ID	Descripción del Objetivo	Prioridad
OBJ1	Implementar un Sistema Inteligente que permita predecir el desempeño de los alumnos a lo largo de la cursada de la materia “Sistemas y Organizaciones” dictada en UTN FRBA.	Alta

Figura 1. Formulario de Objetivos del Proyecto.

INTERESADOS DEL PROYECTO			
Posición	Org/Sector	Rol en el Proyecto	Áreas de Conocimiento
Jefa de Cátedra	UTN- FRBA / Asignatura	Patrocinador	Generalidades, Teoría y Práctica de la Asignatura, y Criterios de Aprobación
Docentes		Interesado	
Ayudantes			Teoría y Práctica de la Asignatura y Repositorios de Datos

Figura 2. Formulario de Interesados del Proyecto.

CRITERIOS DE ÉXITO DEL PROYECTO		
ID	Descripción Criterio	Objetivo
CE1	Predecir a partir de los datos del primer parcial y primer recuperatorio, el desempeño del alumno en el segundo cuatrimestre (incluyendo parcial y recuperatorios).	OBJ1

Figura 3. Formulario de Criterios de Éxito del Proyecto.

DEFINICIÓN DE ALCANCE DEL PROYECTO		
ID	Problemas a Resolver	Objetivo
P1	Identificar las fortalezas y debilidades de los alumnos, de forma de reforzar lo que haga falta durante la cursada.	OBJ1
Problemas Excluidos del Proyecto		
Queda excluido la nota conceptual y la evaluación clase a clase, solo se tomarán en cuenta para el análisis, las notas de parciales y recuperatorios. Asimismo, la planificación anual de la asignatura no es tenida en cuenta para el análisis que se realiza, tomándose solamente como referencia de los temas dictados.		

Figura 4. Formulario de Definición de Alcance de Proyecto.

SUPOSICIONES DEL PROYECTO		
ID	Descripción Suposición	Objetivo
S1	Se tendrá acceso a los ayudantes y docentes sin restricciones, pudiendo realizarse consultas por mail o en forma personal.	OBJ1
S2	Se consideran que los datos son correctos y están completos, teniendo la misma estructura dado que corresponde al mismo docente.	OBJ1

Figura 5. Formulario de Suposiciones del Proyecto.

RESTRICCIONES DEL PROYECTO			
ID	Tipo	Descripción	Objetivo
R1	Datos	No existen datos sobre el progreso del alumno clase a clase.	OBJ1
R2	Datos	No se pueden utilizar las notas de los trabajos prácticos dado que no se encuentran normalizados de la misma forma en todos los cursos.	OBJ1
R3	Datos	No se pueden utilizar los nombres, apellidos y legajos de los alumnos por ser considerados confidenciales.	OBJ1

Figura 6. Formulario de Restricciones del Proyecto.

3.3. Aplicación de la Segunda Fase de Educación de Procesos de Negocio

Las siguientes actividades son descriptas: “Identificar Procesos de Negocio” (sección 3.3.1) y “Relevar Procesos de Negocio” (sección 3.3.2). Asimismo, dado que se desea implementar un Modelo Predictivo en base a los conocimientos de expertos disponibles en la organización, se llevan a cabo las tareas que corresponden a la tercera actividad “Relevar Tareas del Experto” (sección 3.3.3).

3.3.1. Actividad “Identificar Procesos de Negocio”

A partir de las actas de las reuniones realizadas con los interesados del proyecto, del formulario de Objetivos del Proyecto (figura 1), del formulario de Criterios de Éxito del Proyecto (figura 3) y del formulario de Definición del Alcance del Proyecto (figura 4), el Analista Funcional define la actividad del negocio más significativa que se denomina “*Evaluar Exámenes de Alumnos*”. En esta actividad participan los docentes y ayudantes y, tiene por objetivo la corrección de los exámenes de los alumnos para luego pasar los datos a los repositorios destinados para este fin.

3.3.2. Actividad “Relevar Procesos de Negocio”

Teniendo en cuenta la información relevada, asociada al proceso del negocio identificado, el Analista Funcional realiza una nueva reunión con la Jefa de Cátedra y los interesados de la asignatura. De esta manera, se recolecta la información para registrar cómo funciona dicho proceso y cómo es su relación con los repositorios de datos. A partir de lo relevado, se formaliza la información obtenida documentándola en el formulario de procesos de negocio, tal como se muestra en la figura 7.

3.3.3. Actividad “Revelar Tareas del Experto”

A partir de las reuniones realizadas, se detecta que la tarea de determinar si el alumno aprueba, o no, la cursada teniendo en cuentas sus características y el resultado de sus

exámenes depende de muchos conocimientos que se encuentran interiorizados en la mente de los docentes. Esto significa que no existen reglas fijas ni procedimientos estándar para llevarlas a cabo, por lo que no es posible identificar un Proceso de Negocio. Por lo tanto, se trata de una tarea de expertos y se decide llevar a cabo su relevamiento con el fin de obtener los conocimientos aplicados por la Jefa de Cátedra de la asignatura. Para ello, se selecciona como técnica para educir los conocimientos al Análisis de Protocolos, realizando las etapas correspondientes a la técnica según [22]. Una vez efectuados los pasos a partir del protocolo, el Analista Funcional formalizada los conocimientos obtenidos en el formulario tareas del experto, tal como se muestra en la figura 8. Nótese que debido a su tamaño, el Árbol de Descomposición Funcional (Conocimientos Estratégicos) se encuentra disponible en <https://tinyurl.com/yxjwkrj4> y el Mapa de Conocimientos en <https://tinyurl.com/yxbyou6h>

PROCESOS DE NEGOCIO	
P0001 - Evaluar Exámenes de Alumnos	
<i>Actores</i>	Docente; Ayudante
<i>Pre-Condiciones</i>	Los alumnos resolvieron los exámenes a evaluar y ya fueron repartidos entre los docentes y ayudantes para su corrección.
<i>Post-Condiciones</i>	Los exámenes han sido corregidos y las notas pasadas a la planilla del curso.
Flujo Normal	
<ol style="list-style-type: none"> 1. El ayudante corrige la parte práctica de los parciales. 2. El docente corrige la parte teórica de los parciales. 3. El ayudante se reúne con el docente a fin de consultar las dudas que surgieron de la corrección. 4. El docente define la nota final del parcial. 5. El ayudante se encarga de pasar las notas a la planilla correspondiente. 	
Flujo Alternativo	
<ol style="list-style-type: none"> 1. En el caso de que algún examen se haya corregido en forma incorrecta luego de que el alumno lo analice, el docente y ayudante se vuelven a reunir a fin de corregirlo nuevamente y definir la nota final. 2. Una vez decidida la nota, el ayudante se encarga de pasarla a la planilla correspondiente. 	

Figura 7. Formulario de Procesos de Negocio.

TAREAS DEL EXPERTO		
TE001 – Evaluar Cursada del Alumno		
Conocimientos Fáclicos – Tabla Concepto – Característica – Valor		
<i>Conceptos</i>	<i>Características</i>	<i>Valores</i>
<i>Alumno</i>	<i>Legajo Alumno</i>	<i>Alfanumérico</i>
	<i>Apellido Alumno</i>	<i>Alfanumérico</i>
	<i>Nombre Alumno</i>	<i>Alfanumérico</i>
<i>Evaluación</i>	<i>Tipo Evaluación</i>	<i>Alfanumérico</i>
	<i>Nota Teoría</i>	<i>Alfanumérico</i>
	<i>Nota Práctica</i>	<i>Alfanumérico</i>
Conocimientos Tácticos – Pseudo-Reglas		
<ul style="list-style-type: none"> • Si el alumno aprueba la teoría del primer parcial o la teoría del primer recuperatorio del primer parcial o la teoría del segundo recuperatorio del primer parcial entonces el alumno aprueba la teoría del primer cuatrimestre. • Si el alumno aprueba la práctica del primer parcial o la práctica del primer recuperatorio del primer parcial o la práctica del segundo recuperatorio del primer parcial entonces el alumno aprueba la práctica del primer cuatrimestre. • Si el alumno aprueba la teoría del primer cuatrimestre y la práctica del primer cuatrimestre y la teoría del segundo cuatrimestre y la práctica del segundo cuatrimestre entonces aprueba la cursada. • Si el alumno no aprueba la teoría del primer parcial entonces es probable que no apruebe la teoría del segundo parcial. • Si el alumno no aprueba la práctica del primer parcial entonces es probable que no apruebe la práctica del segundo parcial. 		

Figura 8. Formulario de Tareas del Experto.

3.4. Aplicación de la Tercera Fase de Educación de Datos de Procesos de Negocio

Las siguientes actividades son descriptas: “Identificar Repositorios de Datos” (sección 3.4.1) y “Relevar Datos del Negocio” (sección 3.4.2).

3.4.1. Actividad “Identificar Repositorios de Datos”

El Analista Funcional analiza la información de las entrevistas realizadas con los interesados del negocio y el formulario que corresponde a las “Tareas del Experto” (figura 8). Como resultado, detecta que los principales datos a utilizar para la realización del proyecto se encuentran en 5 planillas en formato Excel denominadas “*planificación_<dia cursada>*” (donde *<dia cursada>* corresponde a Lunes, Martes, Jueves, Viernes A y Viernes B), que describen el comportamiento de los alumnos (en cuanto a sus notas de parciales, recuperatorios y trabajos prácticos) durante el cursado. Estas planillas fueron brindadas por la Jefa de Cátedra para proceder a su análisis en las actividades siguientes.

3.4.2. Actividad “Relevar Datos del Negocio”

El Analista Funcional confecciona el formulario de estructuras de datos, que se muestra en la figura 9. Dado que las planillas tienen el mismo formato se presenta una estructura de datos genérica que es válida para todos los repositorios de datos.

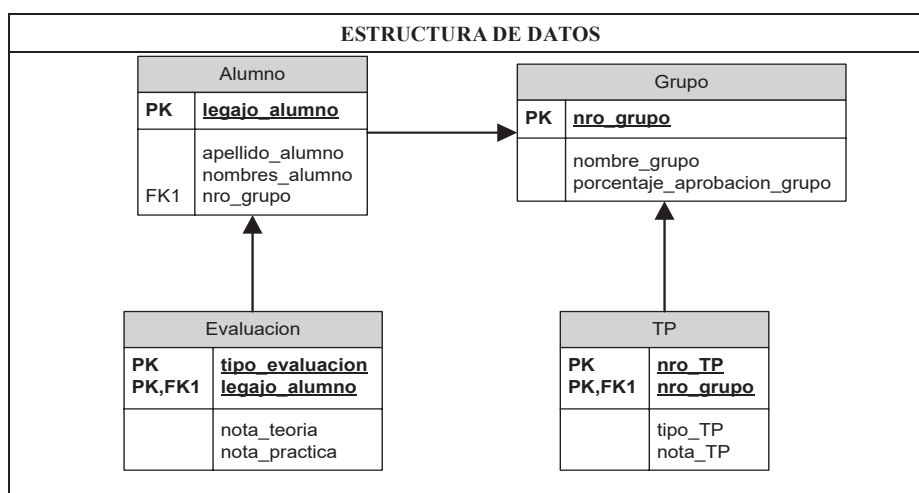


Figura 9. Formulario de Estructura de Datos.

3.5. Aplicación de la Cuarta Fase de Conceptualización de los Datos del Negocio

En esta fase se busca definir si los datos que se utilizarán en el Modelo Predictivo son representativos. En este caso de estudio se realizan dos ciclos de esta fase que se describen a continuación:

- a) En primera instancia se analiza una *Versión Inicial de los Datos* del negocio más general. Para ello, el Analista Funcional trabaja con la información relevada y las 5 planillas recolectadas en la fase anterior. Aunque las mismas constan de tres solapas (“*Planificación Anual*”, “*TPs*” y “*Parciales*”), para este proyecto se utiliza solamente los datos de los exámenes parciales, ya que los datos referidos a los trabajos

prácticos y los grupos de alumnos no se encuentran normalizados para todos los cursos. Asimismo, tampoco se pueden utilizar los nombres, apellidos y legajos de los alumnos por ser considerados confidenciales. Ambas situaciones figuran en el formulario de Restricciones del Proyecto (figura 6).

Luego, se integran las planillas en una sola que incluye 14 atributos correspondientes al desempeño de los alumnos en teoría y práctica de los exámenes (parciales y recuperatorios) así como el día del curso y el resultado de la cursada. De estos datos, se toma una muestra de 75 registros con el objetivo de analizar si son representativos del negocio para construir el Modelo Predictivo. Debido a limitaciones de espacio, todas las tareas aplicadas sobre el conjunto de datos se especifican en [23]. A partir de las conclusiones obtenidas se detecta que esta versión inicial de los datos presenta sesgos que generan diferencias con las relaciones preconcebidas por el docente de los cursos. Es por esto que, se decide generar una nueva versión que se describe a continuación.

- b) En el segundo ciclo, se evalúa una *Versión Extendida de los Datos* la cual incluye mayor cantidad de ejemplos y atributos. Para extender los datos se procede a detallar los resultados de cada tema en cada examen (así, por ejemplo, en la Práctica del primer parcial se indican los resultados para Organigrama y Cursograma). Además, se agregan 51 registros de alumnos, por lo que tienen un total de 126 filas incluyendo 23 atributos. A partir de estos datos se analiza nuevamente si son representativos del negocio para construir el Modelo Predictivo, tal como se detalla en las tareas aplicadas en [23]. Luego del análisis se concluye que la versión extendida es representativa del comportamiento de los alumnos y puede ser utilizada para la construcción del Modelo Predictivo.

3.6. Aplicación de la Quinta Fase de Especificación Inicial del Sistema Inteligente

Las siguientes actividades son descriptas: “Seleccionar el Tipo de Sistema Inteligente” (sección 3.6.1) y “Definir Topología Inicial del Sistema Inteligente” (sección 3.6.2). Asimismo, el proceso completo se encuentra especificado en [24].

3.6.1. Actividad “Seleccionar el Tipo de Sistema Inteligente”

El Analista Funcional responde las preguntas asociadas a cada característica a partir de las actas de reuniones realizadas en fases anteriores, formalizándolas en la tabla 1. Una vez asignados los valores lingüísticos correspondientes para cada característica definida en la tabla 1, el Analista Funcional obtiene los valores correspondientes para cada arquitectura y procede a seleccionar la mejor arquitectura para el proyecto. Las operaciones realizadas para esta actividad se presentan en la planilla de cálculo disponible en [25]. Tal como se muestra en la figura 10, para este caso la arquitectura seleccionada son las *Redes Bayesianas*. Esta selección es justificada, por un lado, por los datos disponibles que incluyen una cantidad suficiente de ejemplos para realizar el entrenamiento y validación de la red pudiendo definir rangos de valores para cada uno. Por otro lado, además se dispone de expertos del dominio (Jefe de Cátedra y Docentes) que pueden aportar su conocimiento y participar en el proyecto. Por lo tanto, se podrán realizar comparaciones entre las predicciones para diferentes escenarios y conocer cómo se obtuvieron los resultados así como realizar ajustes manuales entre los estados de la red que se defina.

Tabla 1. Características Evaluadas para definir la Arquitectura más adecuada.

ID	Pregunta Asociada a la Característica	Valor
D1	¿Cuánta confianza se tiene que los datos son representativos?	<i>mucho</i>
D2	¿En qué medida se pueden considerar a los datos como complejos y con una relación no lineal entre sus atributos?	<i>poco</i>
D3	¿Cuánta cantidad de ejemplos incluyen los datos?	<i>regular</i>
D4	¿Cuál es la proporción de datos con valores numéricos continuos con respecto a discretos?	<i>regular</i>
R1	¿Cuán crítica se considera la precisión de la predicción?	<i>mucho</i>
R2	¿En qué medida se desea conocer comparar las predicciones para diferentes escenarios?	<i>mucho</i>
R3	¿Con qué grado de importancia se desea que se puedan explicar cómo se obtuvieron los resultados generados?	<i>mucho</i>
P1	¿Qué tan estable es el problema a resolver?	<i>mucho</i>
P2	¿Cuánta disponibilidad tienen los expertos del dominio para participar en el desarrollo?	<i>todo</i>
P3	¿Qué tan deseable es poder ajustar manualmente la red a partir de los conocimientos sobre los datos?	<i>todo</i>

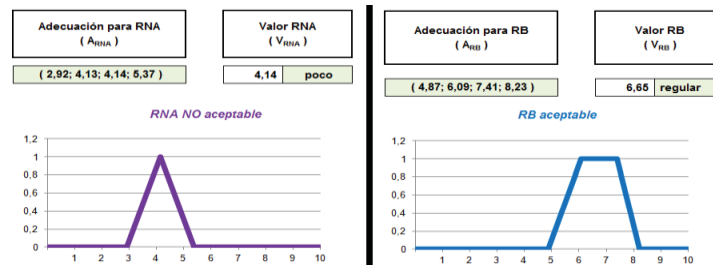


Figura 10. Valores de Adecuación para cada Arquitectura.

3.6.2. Actividad “Definir Topología Inicial del Sistema Inteligente”

Con el tipo de arquitectura ya seleccionado, el Analista Funcional busca determinar un prototipo inicial del modelo propuesto. Para ello, utiliza la herramienta conocida como GeNIe [26], donde decide aplicar una estructura simple donde cada nodo se conecta con el atributo objetivo (CURSADA) y, en el caso de los temas, se vincula el resultado del examen parcial con el primer recuperatorio y éste con el segundo recuperatorio. El motivo de esta relación es que, tal como fue explicado por el experto, se considera que el resultado que tendría un alumno en la teoría del segundo recuperatorio puede estar influenciado por los resultados en el primer recuperatorio y el examen parcial. Con dicha estructura se importan los datos disponibles (correspondientes a la versión extendida obtenida en la fase anterior), para que la herramienta determine las probabilidades correspondientes a cada nodo. Como resultado se obtiene la distribución de probabilidades, tal como se muestra en <https://tinyurl.com/y2c7h7wc>

Aunque la implementación de la Red Bayesiana final queda fuera del alcance de este proceso, con el objetivo de confirmar si esta topología inicial fue entrenada exitosamente, se procede a hacer su validación usando los mismos datos. A pesar de que en un proyecto real no tendría sentido validar un Sistema Inteligente con los mismos datos con que fue entrenado (dado que la precisión así obtenida no es confiable), aquí sólo se busca confirmar que las probabilidades brindadas son representativas de los datos utilizados. Como resultado, la precisión general de esta red es del 94%, con un 100% de precisión para predecir los estudiantes que aprueban el curso, un 87.5% para los estudiantes que no aprueban y un 96.4% para los estudiantes que no terminan la asignatura. Por lo tanto, es posible asegurar que esta topología inicial puede ser usada como un prototipo del Modelo Inteligente para predecir el desempeño de los alumnos de la asignatura.

4. Conclusiones

La implementación de un Modelo Predictivo Inteligente no es una tarea sencilla. Por lo tanto, el proceso propuesto asiste al Ingeniero en los pasos iniciales del proyecto. En este trabajo, se lo ha aplicado en un caso de estudio dentro del ámbito educativo universitario. En la primer fase ha sido posible definir el objetivo que tendrá el Modelo Predictivo. En la segunda fase se ha relevado el proceso de negocio asociado a dicho objetivo. En la tercer fase se han detectado las fuentes de datos, que luego han sido evaluadas en la cuarta fase para identificar un conjunto de datos representativo del comportamiento de los alumnos. Finalmente, en la quinta fase se ha determinado la tecnología más adecuada para construir el Modelo Predictivo.

Referencias

1. De Kohan, N. C. (2008). *Los sesgos cognitivos en la toma de decisiones*. International Journal of Psychological Research, 1(1), 68-73. ISSN 2011-7922.
2. Robbins, S. & Coulter, M. (2010). *Administración. Décima Edición*. Prentice Hall. ISBN: 978-607-442-388-4.
3. Mair, C., Kadoda, G., Lefley, M., Phalp, K., Schofield, C., Shepperd, M. & Webster, S. (2000). *An investigation of machine learning based prediction systems*. Journal of Systems and Software, 53(1), 23-29.
4. García, F. J. M., Martínez, M. A. P. & García, J. S. (2003). *Gestión Estratégica del Conocimiento*. Asociación Universitaria Iberoamericana de Postgrado.
5. Nilsson, N. J. (2014). *Principles of Artificial Intelligence*. Morgan Kaufmann.
6. Russell, S. J., Norvig, P., Davis, E., Russell, S. J. & Russell, S. J. (2010). *Artificial Intelligence: a modern approach* (Vol. 2). Englewood Cliffs: Prentice Hall.
7. Shepperd, M. & Kadoda, G. (2001). *Comparing software prediction techniques using simulation*. IEEE Transactions on Software Engineering, 27(11), 1014-1022.
8. Bontempi, G., Taieb, S. B. & Le Borgne, Y. A. (2013). *Machine learning strategies for time series forecasting*. In Business Intelligence (pp. 62-77). Springer Berlin Heidelberg.
9. Wang, S. C. (2003). *Artificial Neural Network*. In Interdisciplinary Computing in Java Programming (pp. 81-100). Springer US.
10. Wang, L. & Fu, K. (2009). *Artificial Neural Networks*. Wiley Encyclopedia of Computer Science and Engineering, 181-188.
11. Barber, D. (2012). *Bayesian Reasoning and Machine Learning*. The MIT Press.
12. Premchaiswadi, W. (2012) *Bayesian Networks*. Ed. In-Tech.
13. Chatfield, C. (2016). *The analysis of time series: an introduction*. CRC press.
14. De Gooijer, J. G. & Hyndman, R. J. (2006). *25 years of time series forecasting*. International journal of forecasting, 22(3), 443-473.
15. Zhang, G. & Hu, M.Y. (1998). *Neural network forecasting of the British pound/US dollar exchange rate*. Omega, International Journal of Management Science, 26 (4) 495-506.
16. Cohen, P. R. & Feigenbaum, E. A (2014). *The handbook of Artificial Intelligence*. Vol. 3. Butterworth-Heinemann.
17. Acquatela, H. (2006). *La predicción del futuro: desde el oráculo de Delfos hasta la medicina actual*. Gac. méd. Caracas, 114(2), 150-156. ISSN 0367-4762.
18. Rodríguez, M. & Márquez Alegría, M. (2015). *Manejo de problemas y toma de decisiones* (Vol. 8). Editorial El Manual Moderno. SBN: 9789684264670.
19. Arsham, H. (2006). *Tools for Decision Analysis: Analysis of Risky Decisions*.
20. Vegega, C., Pytel, P., Pollo-Cattaneo M. F. (2019). *Aplicación del Proceso de Elicitación de Requerimientos para la Construcción de Modelos Predictivos basados en Sistemas Inteligentes dentro del Ámbito Educativo*. Reporte Técnico GEMIS-TD-2018-05-RT-2019-03. Disponible en <https://tinyurl.com/y2pxnryb>
21. Universidad Tecnológica Nacional – Facultad Regional Buenos Aires (2008). *Programa de la carrera Ingeniería en Sistemas de Información (Plan 2008)*. DISI. Disponible en <https://tinyurl.com/y2kb36xv>
22. Gómez, A., Juristo, N., Montes, C. & Pazos, J. (1997) *Ingeniería del Conocimiento*. Centro de Estudios Ramón Areces. S.A. Madrid.
23. Vegega, C., Pytel, P., Straccia, L., Pollo-Cattaneo, M.F. (2018). *Evaluation of the Bias of Student Performance Data with Assistance of Expert Teacher*. Applied Informatics, Communications in Computer and Information Science book series (CCIS, volume 942). Proceedings of First International Conference (ICAI 2018). Págs 16-31. Online ISSN: 1865-0937.
24. Vegega, C., Pytel, P., Straccia, L., Pollo-Cattaneo, M.F. (2018). *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 (CONAIIISI 2018). Workshop de Aplicaciones Informáticas y de Sistemas de Información. ISSN 2347-0372. Disponible en <https://tinyurl.com/y6nzjhjv>
25. Vegega, C., Pytel, P. & Pollo-Cattaneo M.F. (2017). *Método Evaluador de Arquitectura - Desempeño de Alumnos de Sistemas y Organizaciones*. Disponible en <https://bit.ly/2YwESUM>
26. BayesFusion (2015) *BayesFusion - Home Page for users of GeNe*. BayesFusion LLC. Disponible en <https://www.bayesfusion.com/>