

Caso de Estudio de la aplicación de Procesos de Explotación de Información en Educación

**Deroche, Ariel F.¹, Raus, Nicolás A.¹, Luján, Facundo N.¹,
Vegega, Cinthia¹, Pytel, Pablo^{1,2}, Pollo-Cattaneo, María Florencia¹**

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

*² Grupo Investigación en Sistemas de Información. Departamento Desarrollo Productivo
y Tecnológico. Universidad Nacional de Lanús. Argentina.*

Resumen

La aplicación de la disciplina Explotación de Información en la asignatura “Inteligencia Artificial” ha permitido al equipo docente contar con un conocimiento útil para el dictado de las clases. Aún así, se cuenta con la necesidad de analizar los temas que mayor incidencia tienen en cada parcial de la asignatura. Con este objetivo, se realiza un análisis más profundo aplicando diversos procesos de Explotación de Información en parciales de diferentes años, cuatrimestres y regionales. Finalmente, se presentan los resultados obtenidos con su análisis y conclusiones correspondientes.

Palabras Clave

Explotación de Información. Educación y Tecnología. Análisis de Proceso de Evaluación. Aprendizaje.

Introducción

La Explotación de Información permite, a partir de grandes masas de información, obtener un conocimiento útil que previa interpretación por parte del Experto en el dominio, se puede utilizar para mejorar o solucionar un problema determinado [1]. Los procesos de la Explotación de Información permiten obtener resultados de análisis de la masa de información que los métodos convencionales no logran tales como: los algoritmos Top-Down Induction Deduction Trees (TDIDT), los Mapas Auto-Organizados (SOM) y las Redes Bayesianas [2]. Así, en el caso de los algoritmos TDIDT, ellos permiten el desarrollo de descripciones simbólicas de los datos para diferenciar entre distintas clases [3].

La aplicación de Explotación de Información en la asignatura “Inteligencia Artificial” ha tenido interesantes resultados.

A partir del estudio de los exámenes parciales realizado en [4] para la asignatura perteneciente al 5to año de la carrera “Ingeniería en Sistemas de Información” de la Universidad Tecnológica Nacional (Regional Buenos Aires), se han podido analizar los criterios de aprobación en los exámenes parciales. El conocimiento obtenido ha sido interpretado para intentar encontrar estrategias para mejorar el dictado de la materia.

Se cree que incluir a dichos datos los correspondientes al 2do cuatrimestre del año 2012 y los pertenecientes a la Regional La Plata, puede generar otras conclusiones mucho más útiles, teniendo la posibilidad además de aplicar otros tipos de procesos.

En este contexto, el objetivo de este trabajo es analizar todos los parciales disponibles en conjunto, aplicando diferentes tipos de procesos de Explotación de Información para que la cátedra pueda obtener la información necesaria para tomar las decisiones correspondientes.

Definición del Problema

Los docentes de la cátedra han solicitado un análisis integrador de los parciales que permita determinar los temas principales de la asignatura sin importar el cuatrimestre y regional que cursen. De esta forma se espera obtener conocimiento para la posterior mejora del dictado de las clases y forma de evaluación y reducir la cantidad de exámenes desaprobados. Los porcentajes de desaprobación han aumentado en el último cuatrimestre analizado (2do cuatrimestre del 2012), tal y como se puede

observar en la tabla 1, sobre todo para la Regional La Plata.

Tabla 1. Porcentajes de aprobación de los parciales de la asignatura

Año	Cuat.	Reg.	% Aprobados	% Desaprobados
2011	1ro	BA	71%	29%
2012	1ro	BA	63%	37%
2012	2do	BA	57%	43%
2012	2do	LP	14%	86%
Promedio General			51%	49%

Solución Propuesta

Se cuenta con la totalidad de los exámenes parciales pertenecientes a la Regional Buenos Aires para el 1er cuatrimestre del año 2011 y los dos cuatrimestres del 2012, y los parciales de la Regional La Plata para el 2do cuatrimestre del año 2012.

Para aplicar los procesos correspondientes, primero se realiza un proceso de preparación de carga y normalización de los 351 parciales generando un registro por cada uno.

Por cada registro se generan atributos que representan la cantidad de respuestas correctas e incorrectas por tema como se puede observar en la tabla 2. Además se cuenta con el tema (que codifica la regional y día de cursada) y el resultado del examen (aprobado o desaprobado). Sólo en el caso del tema “Análisis de Protocolos” (AP), que resulta ser un tema clave, se consideran también las respuestas no contestadas. Es decir que para ese caso se establecen los atributos AP_OK (cantidad de respuestas correctas para este tema), AP_MAL (cantidad de respuestas incorrectas para este tema) y AP_N (cantidad de preguntas de este tema no respondidas).

Descubrimiento de Reglas de Comportamiento

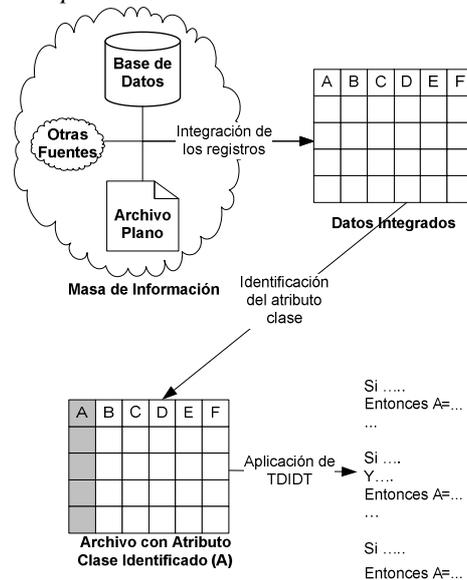
El primer proceso de Explotación de Información utilizado es el de “Descubrimiento de Reglas de Comportamiento” [4]. Este proceso es aplicado cuando se requieren identificar las condiciones que se deben dar para obtener

un resultado (atributo clase) ya conocido en el dominio del problema.

A partir del conjunto de atributos preparados, se aplica el algoritmo de inducción C4.5 perteneciente a la familia TDIDT [5]. El resultado de este proceso es un conjunto de reglas donde el atributo clase aparece en el concluyente y los otros atributos en las condiciones del antecedente. De esta forma, es posible identificar el comportamiento de los datos y las relaciones implícitas entre los atributos.

El esquema de este proceso se puede ver en la figura 1.

Figura 1. Proceso de Descubrimiento de Reglas de Comportamiento.



Como resultado de este proceso se obtienen las siguientes reglas, las cuales han sido agrupadas en cuanto a si influyen en la aprobación o desaprobación del examen:

Reglas que dan como resultado un parcial aprobado:

REGLA A1

SI

TEORIA_MAL	≤ 2	Y
AP_MAL	≤ 1	Y
BUSQ_MAL	= 0	Y
AP_N	= 0	Y
TEMA	≤ 31	

ENTONCES ResultadoD = Aprobado

Tabla 2. Atributos y sus respectivas descripciones

Atributo	Descripción
AP_MAL	Cantidad de respuestas incorrectas para el tema Análisis de Protocolos.
AP_N	Cantidad de preguntas no respondidas para el tema Análisis de Protocolos.
AP_OK	Cantidad de respuestas correctas para el tema Análisis de Protocolos.
BUSQ_MAL	Cantidad de respuestas incorrectas para ejercicios de Métodos de Búsqueda.
BUSQ_OK	Cantidad de respuestas correctas para ejercicios de Métodos de Búsqueda.
DEDNAT_MAL	Cantidad de respuestas incorrectas para ejercicios de Deducción Natural.
DEDNAT_OK	Cantidad de respuestas correctas para ejercicios de Deducción Natural.
EMP_MAL	Cantidad de respuestas incorrectas para ejercicios de Emparrillado.
EMP_OK	Cantidad de respuestas correctas para ejercicios de Emparrillado.
TEORIA_MAL	Cantidad de respuestas incorrectas para preguntas teóricas sobre IA e INCO.
TEORIA_OK	Cantidad de respuestas correctas para preguntas teóricas sobre IA e INCO.
ARQ_MAL	Cantidad de respuestas incorrectas para tema teórico de arquitecturas de sistemas inteligentes
ARQ_OK	Cantidad de respuestas correctas para tema teórico de arquitecturas de sistemas inteligentes
TRAD_MAL	Cantidad de respuestas incorrectas para ejercicios de Traducción Natural.
TRAD_OK	Cantidad de respuestas correctas para ejercicios de Traducción Natural.
TEMA	Indica el curso y tema del parcial. 11= lunes T1 (BA); 12= lunes T2 (BA); 21= miércoles R1 (BA); 22= miércoles T2 (BA); 31= sábado T1 (LP); 32= sábado T2 (LP).
ResultadoD	“Aprobado” o “Desaprobado”.

REGLA A2

SI

TRAD_OK	≥ 1	Y
BUSQ_OK	≥ 2	Y
DEDNAT_OK	$= 1$	Y
AP_MAL	≤ 1	Y
EMP_OK	≤ 2	Y
BUSQ_MAL	≥ 1	Y
AP_N	$= 0$	Y
TEMA	≤ 31	

ENTONCES ResultadoD = Aprobado

REGLA A3

SI

ARQ_MAL	≤ 1	Y
ARQ_OK	≥ 2	Y
TEORIA_MAL	≤ 2	Y
EMP_OK	≥ 3	Y
BUSQ_MAL	≥ 1	Y
AP_N	$= 0$	Y
TEMA	≤ 31	

ENTONCES ResultadoD = Aprobado

REGLA A4

SI

ARQ_OK	≥ 2	Y
TEORIA_MAL	≤ 2	Y
EMP_OK	≥ 3	Y
BUSQ_MAL	$= 1$	Y
AP_N	$= 0$	Y
TEMA	≤ 31	

ENTONCES ResultadoD = Aprobado

REGLA A5

SI

TRAD_MAL	$= 0$	Y
TEORIA_MAL	≥ 3	Y
AP_MAL	≤ 1	Y
BUSQ_MAL	$= 0$	Y
AP_N	$= 0$	Y
TEMA	≤ 31	

ENTONCES ResultadoD = Aprobado

REGLA A6

SI

TEORIA_OK	≥ 3	Y
DEDNAT_MAL	$= 0$	Y
EMP_OK	≤ 2	Y
BUSQ_MAL	≥ 1	Y
AP_N	$= 0$	Y
TEMA	≤ 31	

ENTONCES ResultadoD = Aprobado

REGLA A7

SI

ARQ_OK	≥ 2	Y
EMP_OK	≥ 3	Y
AP_MAL	≥ 2	Y
BUSQ_MAL	$= 0$	Y
AP_N	$= 0$	Y
TEMA	≤ 31	

ENTONCES ResultadoD = Aprobado

De estas reglas obtenidas, se presentan las siguientes conclusiones:

- La totalidad de las reglas de aprobación aplican para todos los cursos de la Regional Buenos Aires (Lunes y Miércoles) y sólo para el tema 1 de los Sábados (Regional La Plata). Esto está indicado por la condición $TEMA \leq 31$.
- Por la Regla A1, se puede decir que si se contesta mal hasta una pregunta del tema “Análisis de Protocolo” y hasta dos del tema “Teoría”, sin tener preguntas mal respondidas del tema “Métodos de Búsqueda”, el parcial resulta aprobado.
- Las Regla A2 y A5 también otorgan un grado de importancia al tema “Análisis de Protocolo”, teniendo en cuenta que a pesar de tener una o más preguntas mal respondidas del tema “Métodos de Búsqueda” (en el caso de la Regla A2) y tres o más preguntas mal del tema “Teoría” (Regla A5), contando con una sola pregunta mal de “Análisis de Protocolo” y el resto bien (además de no contar con preguntas mal de otros temas como “Traducción Natural” y “Métodos de Búsqueda”, entre otros), se aprueba el parcial.
- Por lo contrario, la Regla A7 otorga un peso importante a los temas “Arquitectura de Sistemas Inteligentes” y “Emparrillado”, siendo los temas con mayor cantidad de preguntas por parcial. Esto confirma lo previamente detectado en el estudio anterior [4].
- Las Reglas A3 y A4 son muy similares respecto a los temas “Emparrillado” (tres o más preguntas bien respondidas), “Teoría” (hasta dos preguntas mal

respondidas), “Arquitecturas de Sistemas Inteligentes” (dos o más preguntas bien respondidas) y “Análisis de Protocolos” (tener todas las preguntas respondidas).

- Por último, la Regla A6 permite visualizar un grado de importancia otorgado a la “Teoría” ya que contestar bien tres o más preguntas de dicho tema, no tener preguntas mal respondidas de “Deducción Natural” y tener todas las preguntas respondidas de “Análisis de Protocolos”, a pesar de contar con hasta dos preguntas bien respondidas del tema “Emparrillado” y con una o más preguntas mal respondidas de “Métodos de Búsqueda”, se aprueba el parcial.

Reglas que dan como resultado un parcial desaprobado:

REGLA D1

SI

$$TEMA \geq 32$$

ENTONCES ResultadoD = Desaprobado

REGLA D2

SI

$$AP_N \geq 1 \quad Y$$

$$TEMA \leq 31$$

ENTONCES ResultadoD = Desaprobado

REGLA D3

SI

$$EMP_OK \leq 2 \quad Y$$

$$AP_MAL \geq 2 \quad Y$$

$$BUSQ_MAL = 0 \quad Y$$

$$AP_N = 0 \quad Y$$

$$TEMA \leq 31$$

ENTONCES ResultadoD = Desaprobado

REGLA D4

SI

$$ARQ_OK \leq 1 \quad Y$$

$$EMP_OK \geq 3 \quad Y$$

$$AP_MAL \geq 2 \quad Y$$

$$BUSQ_MAL = 0 \quad Y$$

$$AP_N = 0 \quad Y$$

$$TEMA \leq 31$$

ENTONCES ResultadoD = Desaprobado

REGLA D5**SI**

TRAD_MAL	≥ 1	Y
TEORIA_MAL	≥ 3	Y
AP_MAL	≤ 1	Y
BUSQ_MAL	$= 0$	Y
AP_N	$= 0$	Y
TEMA	≤ 31	

ENTONCES ResultadoD = Desaprobado**REGLA D6****SI**

ARQ_OK	≤ 1	Y
TEORIA_MAL	≤ 2	Y
EMP_OK	≥ 3	Y
BUSQ_MAL	≥ 1	Y
AP_N	$= 0$	Y
TEMA	≤ 31	

ENTONCES ResultadoD = Desaprobado**REGLA D7****SI**

TEORIA_OK	≤ 2	Y
DEDNAT_MAL	$= 0$	Y
EMP_OK	≤ 2	Y
BUSQ_MAL	≥ 1	Y
AP_N	$= 0$	Y
TEMA	≤ 31	

ENTONCES ResultadoD = Desaprobado**REGLA D8****SI**

ARQ_MAL	$= 2$	Y
BUSQ_MAL	≥ 2	Y
TEORIA_MAL	≤ 2	Y
EMP_OK	≥ 3	Y
AP_N	$= 0$	Y
TEMA	≤ 31	

ENTONCES ResultadoD = Desaprobado**REGLA D9****SI**

AP_MAL	≥ 2	Y
DEDNAT_MAL	≥ 1	Y
EMP_OK	≤ 2	Y
BUSQ_MAL	≥ 1	Y
AP_N	$= 0$	Y
TEMA	≤ 31	

ENTONCES ResultadoD = Desaprobado**REGLA D10****SI**

DEDNAT_OK	$= 0$	Y
AP_MAL	≤ 1	Y
EMP_OK	≤ 2	Y
BUSQ_MAL	≥ 1	Y
AP_N	$= 0$	Y
TEMA	≤ 31	

ENTONCES ResultadoD = Desaprobado**REGLA D11****SI**

TRAD_OK	$= 0$	Y
BUSQ_OK	≥ 2	Y
DEDNAT_OK	$= 1$	Y
AP_MAL	≤ 1	Y
DEDNAT_MAL	$= 1$	Y
EMP_OK	≤ 2	Y
BUSQ_MAL	≥ 1	Y
AP_N	$= 0$	Y
TEMA	≤ 31	

ENTONCES ResultadoD = Desaprobado**REGLA D12****SI**

BUSQ_OK	≤ 1	Y
DEDNAT_OK	$= 1$	Y
AP_MAL	≤ 1	Y
DEDNAT_MAL	$= 1$	Y
EMP_OK	≤ 2	Y
AP_N	$= 0$	Y
TEMA	≤ 31	

ENTONCES ResultadoD = Desaprobado**REGLA D13****SI**

TEORIA_MAL	≥ 3	Y
EMP_OK	≥ 3	Y
BUSQ_MAL	≥ 1	Y
AP_N	$= 0$	Y
TEMA	≤ 31	

ENTONCES ResultadoD = Desaprobado

Según se puede observar en estas reglas obtenidas:

- La Regla D1 señala que todos los parciales del tema 2 del día Sábado (correspondientes a La Plata) desaprobaban (TEMA ≥ 32).
- La Regla D2 vuelve a mostrar la importancia del tema “Análisis de Protocolo”, ya que teniendo una o más

preguntas sin contestar del mismo, implica la desaprobación del examen.

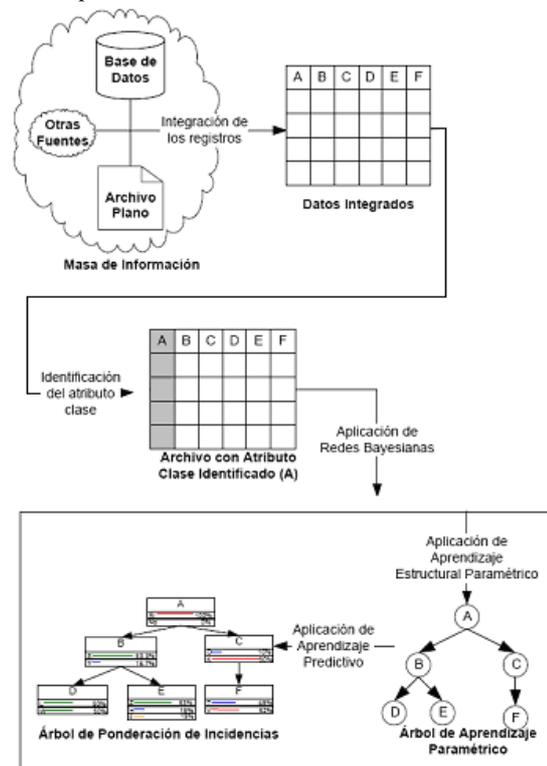
- Las Reglas D3, D4 y D9 son similares a la anterior. También se refieren a “Análisis de Protocolo”, indicando que el contestar dos o más preguntas mal de ese tema implica desaprobar a pesar de contar con otras preguntas bien respondidas como “Emparrillado”.
- Cuando la cantidad de preguntas mal respondidas de “Teoría” supera en tres (que muchas veces representa la totalidad en cantidad de preguntas por parcial), resulta ser más importante que el tema “Análisis de Protocolo”, como se puede observar en las Reglas D5 y D13.
- En las Reglas D6 y D8 se puede observar que presentan atributos similares: contar con hasta por lo menos dos preguntas mal respondidas de “Teoría”, tres o más preguntas bien respondidas de “Emparrillado” y contestar todas las preguntas de “Análisis de Protocolos”. La diferencia entre ambos se encuentra en la cantidad de preguntas bien y mal respondidas de los temas “Arquitecturas de Sistemas Inteligentes” y “Métodos de Búsqueda”.
- Las Reglas D10, D11 y D12 se caracterizan por contar con todas las preguntas de “Análisis de Protocolo” respondidas y a su vez contar con hasta una mal respondida de este tema. Esto significa que en estas reglas existen por lo menos dos preguntas bien respondidas de dicho tema. Esto parecería contradecir lo indicado anteriormente debido al grado de importancia que se le ha otorgado al tema. Pero las reglas cuentan con hasta dos preguntas bien respondidas del tema “Emparrillado” lo que equivale a contar con por lo menos tres preguntas mal respondidas de dicho tema. Esto explica, en parte, el resultado del parcial. Luego existe la presencia de preguntas mal respondidas de los temas “Métodos de Búsqueda” y “Deducción Natural”

Ponderación de Atributos

Como también se desea identificar cuáles son los atributos de mayor incidencia u ocurrencia al momento de la aprobación o desaprobación de un examen parcial, se aplica el proceso de “Ponderación de Interdependencia de Atributos” [2].

De esta manera, mediante la utilización del algoritmo Naïve Bayes [6] es posible identificar el grado de influencia de los atributos que forman parte del dominio del problema sobre un atributo clase previamente seleccionado. En la figura 2 se puede observar el proceso mencionado.

Figura 2. Proceso de Ponderación de Interdependencia de Atributos.



Considerando solamente los principales atributos (previamente identificados a partir de las reglas obtenidas), se realiza un análisis de la probabilidad condicional para cada atributo en forma independiente.

• Regional:

En la figura 3 se puede observar una diferencia importante en la cantidad de aprobados según la regional donde se cursó

la asignatura. Mientras que el 57% de los alumnos de la Regional Buenos Aires aprueba, en la Regional La Plata el porcentaje es sólo del 14%.

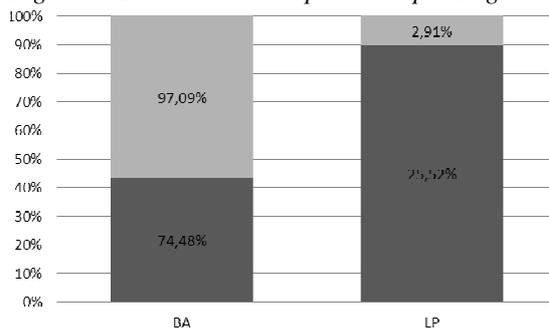
Por otro lado, en la tabla 3 se puede observar que sobre el total de los exámenes analizados, aproximadamente el 97% de los aprobados corresponden a Buenos Aires mientras que el 3% restante pertenecen a La Plata.

Esto confirma lo previamente conocido que se había identificado en la tabla 1.

Tabla 3. Ponderación del atributo Regional

Regional	Desaprobado	Aprobado
BA	74,48%	97,09%
LP	25,52%	2,91%

Figura 3. Distribución de aprobados por Regional



Es importante aclarar que el color gris oscuro corresponde al porcentaje de desaprobados mientras el color gris claro corresponde a los aprobados.

■ Aprobado

Es decir: ■ Desaprobado

- Análisis de Protocolos:

Siendo este tema uno de los más influyentes a la hora de aprobar o no un examen parcial (según las reglas obtenidas en la sección anterior), se puede observar en la tabla 4 como el 50% de los aprobados no tienen preguntas mal respondidas de dicho tema, y como casi aproximadamente el 82% de los desaprobados posee hasta dos preguntas mal respondidas.

Esto confirma que a mayor cantidad de preguntas mal respondidas de este tema, la probabilidad de desaprobación aumenta.

En la figura 4 se puede apreciar con mayor detalle que el responder mal todas las preguntas de este tema conlleva a tener una probabilidad de sólo el 12% de aprobar el parcial (a pesar de los resultados de los otros temas).

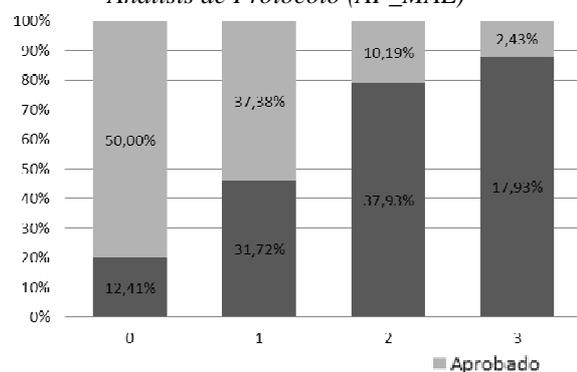
Por otro lado, según se observa en la tabla 5, aproximadamente el 44% de los exámenes aprobados han contado con tres preguntas bien respondidas del tema.

En relación con lo mencionado en el párrafo anterior, la figura 5 permite visualizar con más detalles como el tener tres o más preguntas bien respondidas de dicho tema equivale a tener un 80% de probabilidades de aprobación.

Tabla 4. Ponderación de aprobados por cantidad de preguntas mal respondidas del tema Análisis de Protocolo (AP_MAL)

Cantidad de AP_MAL	Desaprobado	Aprobado
0	12,41%	50,00%
1	31,72%	37,38%
2	37,93%	10,19%
3	17,93%	2,43%

Figura 4. Distribución de aprobados por cantidad de preguntas mal respondidas del tema Análisis de Protocolo (AP_MAL)



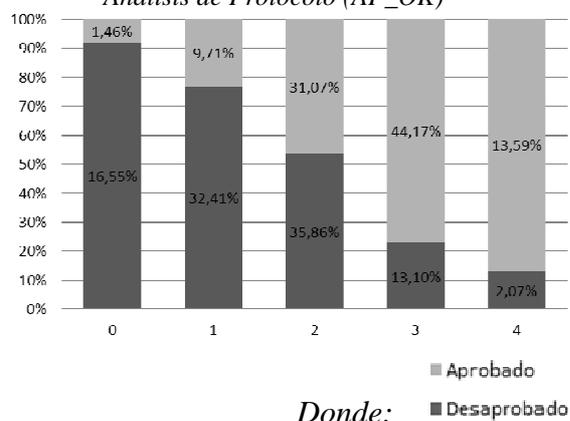
Donde: ■ Aprobado

■ Desaprobado

Tabla 5. Ponderación de aprobados por cantidad de preguntas bien respondidas del tema *Análisis de Protocolo (AP_OK)*

Cantidad de AP_OK	Desaprobado	Aprobado
0	16,55%	1,46%
1	32,41%	9,71%
2	35,86%	31,07%
3	13,10%	44,17%
4	2,07%	13,59%

Figura 5. Distribución de aprobados por cantidad de preguntas bien respondidas del tema *Análisis de Protocolo (AP_OK)*



- Métodos de Búsqueda:

Contar con dos o más preguntas bien respondidas de este tema incrementa las probabilidades de aprobación (más del 60%) como se observa en la tabla y figura 6. Pero este tema no influye mucho para la aprobación general: se puede observar esto en el bajo porcentaje de alumnos que aprobaron (casi el 2%) sin contar con preguntas bien respondidas de este tema. Y que desaprobaron contando con todas las preguntas bien respondidas (el 3,45%).

- Deducción Natural:

El tema "Deducción Natural" presenta situación similar. Existe una probabilidad de aprobación de más del 60%, contando con dos o más preguntas bien respondidas, tal y como se puede visualizar en la tabla y figura 7. Pero también se debe destaca el

bajo porcentaje de alumnos que aprueban sin contar con preguntas bien respondidas del tema (casi el 6%) y que desaprobaran habiendo respondido bien todas las preguntas presentes (casi el 3%).

Tabla 6. Ponderación de aprobados por cantidad de preguntas bien respondidas del tema *Métodos de Búsqueda (BUSQ_OK)*

Cant. de BUSQ_OK	Desaprobado	Aprobado
0	13,79%	1,94%
1	47,59%	12,62%
2	35,17%	58,74%
3	3,45%	26,70%

Figura 6. Distribución de aprobados por cantidad de preguntas bien respondidas del tema *Métodos de Búsqueda (BUSQ_OK)*

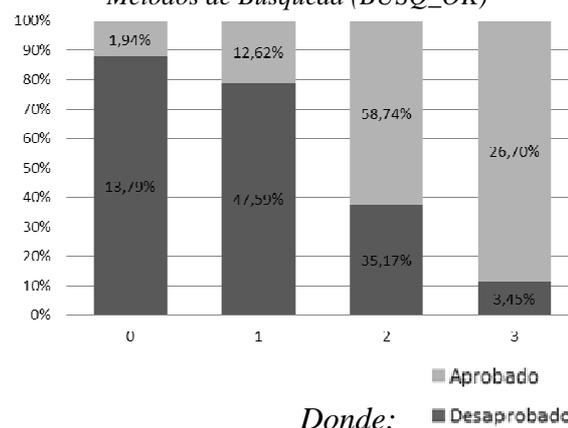
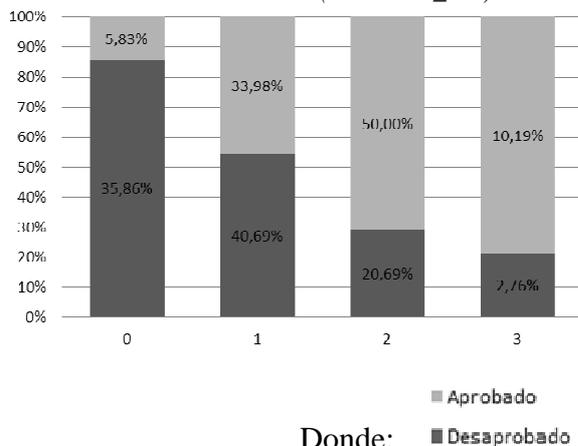


Tabla 7. Ponderación de aprobados por cantidad de preguntas bien respondidas del tema *Deducción Natural (DEDNAT_OK)*

Cant. de DEDNAT_OK	Desaprobado	Aprobado
0	35,86%	5,83%
1	40,69%	33,98%
2	20,69%	50,00%
3	2,76%	10,19%

Figura 7. Distribución de aprobados por cantidad de preguntas bien respondidas del tema Deducción Natural (DEDNAT_OK)



• Emparrillado:

Este tema es el que cuenta, en promedio, con la mayor cantidad de preguntas por parcial.

Según se puede observar en la tabla 8, los porcentajes de aprobados (5,34%) y desaprobados (5,52%) son similares cuando se cuenta con las cinco preguntas bien respondidas del tema. Por otro lado, las probabilidades de aprobación del examen cuando no se cuenta con preguntas bien respondidas del tema es muy baja (sólo el 0,49%).

La figura 8 presenta un comportamiento irregular cuando se cuenta con 4 o más preguntas bien respondidas del tema: la probabilidad de aprobación asciende a 90% cuando se cuenta con cuatro bien respondidas pero la misma desciende al 50% cuando se cuenta con cinco bien respondidas. Este comportamiento extraño se genera debido que, a modo de prueba, en el 1er cuatrimestre del año 2011 se decidió incluir seis preguntas de este tema en los parciales del día miércoles. Luego, en los siguientes cuatrimestres, se bajo la cantidad de preguntas de este tema a cuatro, aumentando en dos la cantidad de preguntas del tema “Teoría”.

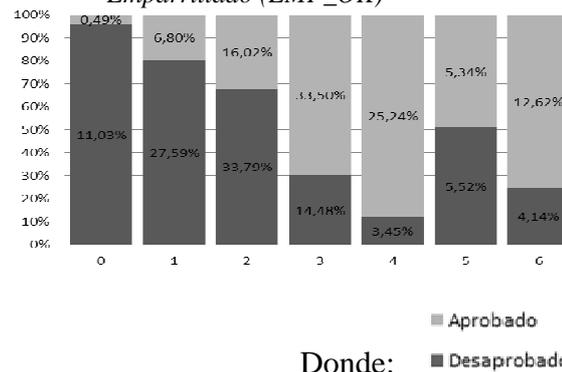
Como se puede observar en la tabla 8, la cantidad de exámenes con cinco o seis preguntas de este tema representa sólo el

9,66% del total de exámenes desaprobados y el 17,96% de los aprobados.

Tabla 8. Ponderación de aprobados por cantidad de preguntas bien respondidas del tema Emparrillado (EMP_OK)

Cant. de EMP_OK	Desaprobado	Aprobado
0	11,03%	0,49%
1	27,59%	6,80%
2	33,79%	16,02%
3	14,48%	33,50%
4	3,45%	25,24%
5	5,52%	5,34%
6	4,14%	12,62%

Figura 8. Distribución de aprobados por cantidad de preguntas bien respondidas del tema Emparrillado (EMP_OK)



• Teoría:

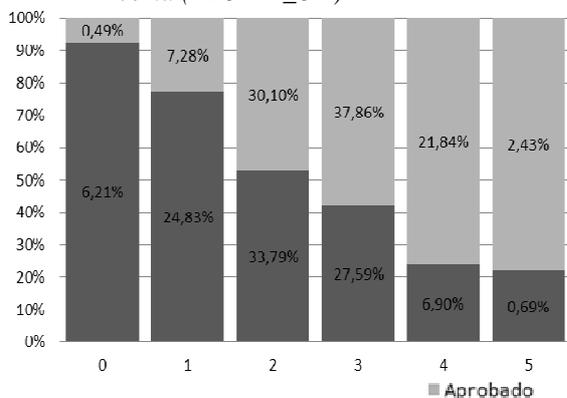
Tener tres o más preguntas bien respondidas de “Teoría” genera una gran posibilidad de aprobar el examen (ver figura 9).

Además se puede observar en la tabla 9 que en muy pocos parciales se han utilizado cinco preguntas de este tema (3,12% entre desaprobados y aprobados). A su vez, la cantidad de aprobados sin tener preguntas bien respondidas del tema resulta muy baja (0,49%) lo que equivale a aproximadamente 2 parciales del total de 351.

Tabla 9. Ponderación de aprobados por cantidad de preguntas bien respondidas del tema Teoría (TEORIA_OK)

Cant. de TEORIA_OK	Desaprobado	Aprobado
0	6,21%	0,49%
1	24,83%	7,28%
2	33,79%	30,10%
3	27,59%	37,86%
4	6,90%	21,84%
5	0,69%	2,43%

Figura 9. Distribución de aprobados por cantidad de preguntas bien respondidas del tema Teoría (TEORIA_OK)



Donde: ■ Desaprobado

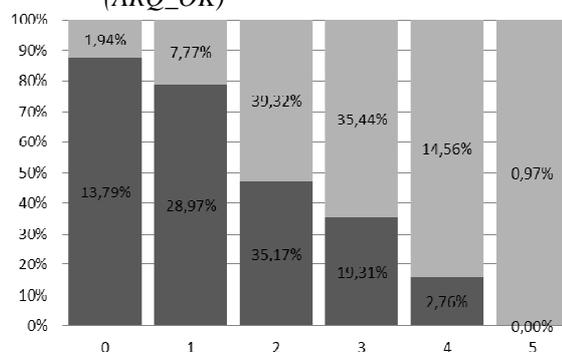
• Arquitecturas de Sistemas Inteligentes:
Tal y como se puede observar en la figura 10, contar con dos o más preguntas bien respondidas de este tema genera buenas probabilidades de aprobación (más del 50%). Las probabilidades aumentan cuando se cuenta con cuatro o más preguntas bien respondidas.

Pero es importante comentar la baja cantidad de parciales que cuentan con 5 preguntas bien respondidas del tema en cuestión (menos del 1% del total de los exámenes). De estos exámenes, todos han sido aprobados (ver tabla 10).

Tabla 10. Ponderación de aprobados por cantidad de preguntas bien respondidas del tema Arquitectura de Sistemas Inteligentes (ARQ_OK)

Cant. de ARQ_OK	Desaprobado	Aprobado
0	13,79%	1,94%
1	28,97%	7,77%
2	35,17%	39,32%
3	19,31%	35,44%
4	2,76%	14,56%
5	0,00%	0,97%

Figura 10. Distribución de aprobados por cantidad de preguntas bien respondidas del tema Arquitectura de Sistemas Inteligentes (ARQ_OK)



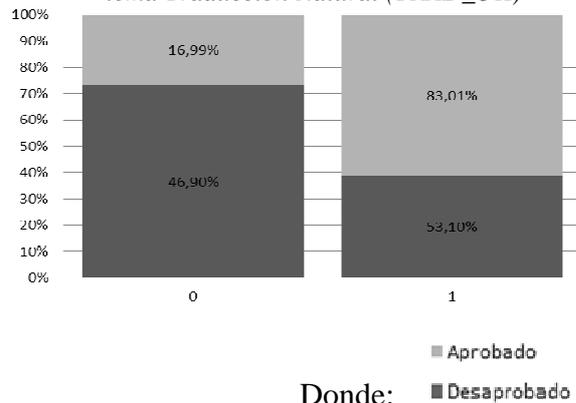
Donde: ■ Desaprobado

• Traducción Natural:
Si bien este tema no resulta relevante a la hora de aprobar el parcial, contar con una pregunta bien de este tema posee asociado un 60% de probabilidades de aprobación, tal y como se observa en la figura 11. Por otro lado, el 70% de los alumnos que desaprueban no poseen preguntas bien respondidas de este tema. Observando con detalle la tabla 11, se nota que el 83% de los aprobados cuenta con esta pregunta bien respondida.

Tabla 11. Ponderación de aprobados por cantidad de preguntas bien respondidas del tema Traducción Natural (TRAD_OK)

Cant. de TRAD_OK	Desaprobado	Aprobado
0	46,90%	16,99%
1	53,10%	83,01%

Figura 11. Distribución de aprobados por cantidad de preguntas bien respondidas del tema Traducción Natural (TRAD_OK)



Análisis de los Resultados

En el estudio previo realizado en [4] se habían presentado y analizado un conjunto de 20 reglas. En este nuevo estudio, y utilizando una cantidad mayor de parciales mucho más heterogénea (por incluir otra regional donde se dicta la misma asignatura), se han obtenido conclusiones similares.

Si bien el tema “Análisis de Protocolos” resulta tener una importancia similar respecto al análisis anterior, se puede observar, según las nuevas reglas, que los temas que poseen mayor cantidad de preguntas aumentan su influencia sobre la aprobación del parcial. Este es el caso de los temas “Arquitectura de Sistemas Inteligentes” y “Emparrillado”. El análisis realizado con la ponderación de la interdependencia del tema “Análisis de Protocolo” permite visualizar el alto grado de probabilidades de aprobación cuando se cuenta con tres preguntas bien respondidas de dicho tema, mientras que contar con tres preguntas mal respondidas, las probabilidades de aprobación se reducen al 12%.

Por otro lado, el tema “Deducción Natural” presenta en ciertos aspectos un comportamiento similar al tema “Métodos de Búsqueda”:

- o primero, la cantidad promedio de preguntas por parcial de ambos temas resultan similares.

- o segundo, por la probabilidad de aprobación (más del 60%) contando con dos o más preguntas bien respondidas.
- o tercero, por el bajo porcentaje de alumnos que aprueban sin contar con preguntas bien respondidas del tema (5,83%) y que desaprueban habiendo respondido bien todas las preguntas presentes (2,76%).

Aun así, una de las diferencias que podemos mencionar respecto a estos temas se presenta cuando hay una única pregunta bien respondida de ambos temas: en el caso de “Deducción Natural”, la probabilidad de aprobación es de 20% aproximadamente, mientras para “Métodos de Búsqueda” dicha probabilidad aumenta hasta un 45%. Entonces, cuando todas las preguntas están bien respondidas, el tema “Métodos de Búsqueda” resulta más influyente que “Deducción Natural”. Pero cuando hay pocas preguntas bien respondidas, éste último resulta más importante.

Conclusión

Los procesos de Explotación de Información utilizados en este trabajo han permitido obtener un conocimiento amplio y detallado de los temas que inciden en la aprobación o no de los exámenes.

La ponderación de atributos ha proporcionado el grado de incidencia de cada tema evaluado para facilitar la interpretación por el equipo docente. Contar con esta información resulta muy útil para determinar la cantidad de preguntas que se deben incluir en el examen dependiendo de las dificultades que hayan tenido los alumnos durante las clases. En otras palabras, es posible ajustar la cantidad de preguntas en función del grado de participación por parte de los alumnos para así revertir el déficit de aprobación. Si bien no es información novedosa para la cátedra, se ha comprobado que los temas que poseen mayor cantidad de preguntas por examen obtienen una mayor relevancia.

Por otra parte, las reglas obtenidas han permitido confirmar los resultados

obtenidos en el trabajo anterior por ser estos muy similares.

Todos estos resultados obtenidos (prometedores) plantean un interrogante para ser analizado en el futuro: ¿de qué manera el equipo docente debería ajustar la planificación del dictado de los diferentes temas para mejorar la comprensión de los alumnos y generar así un mayor porcentaje de aprobación en los exámenes parciales?

Referencias

1. Schiefer, J., Jeng, J., Kapoor, S. & Chowdhary, P. (2004). Process Information Factory: A Data Management Approach for Enhancing Business Process Intelligence. Proceedings 2004 IEEE International Conference on E-Commerce Technology. Pág. 162-169.
2. Britos, P., García-Martínez, R. (2009). Propuesta de Procesos de Explotación de Información. Proceedings XV Congreso Argentino de Ciencias de la Computación. Workshop de Base de Datos y Minería de Datos. Págs. 1041-1050. ISBN 978-897-24068-4-1.
3. Kuna, H., García Martínez, R., Villatoro, F. (2009). Identificación de Causales de Abandono de Estudios Universitarios. Uso de Procesos de Explotación de Información. Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología 5: 39-44.
4. Deroche, A., Pytel, P., Pollo-Cattaneo, F. (2013). Propuesta de mejora en asignatura de grado mediante Explotación de Información. Proceedings VIII Congreso de Tecnología en Educación y Educación en Tecnología. Artículo ID 5492. ISBN 978-987-1676-04-0.
5. Quinlan J.. (1990) Learning Logic Definitions from Relations. Machine Learning. 5:239-266.
6. Britos, P., Felgaer, P., García-Martínez, R. (2008). Bayesian Networks Optimization Based on Induction Learning Techniques. In Artificial Intelligence in Theory and Practice II, ed. M. Bramer, (Boston: Springer), en prensa.

Datos de Contacto:

Ariel Deroche. *GEMIS. Universidad Tecnológica Nacional, Regional Buenos Aires.* arielderoche@gmail.com.

Nicolás Raus. *GEMIS. Universidad Tecnológica Nacional, Regional Buenos Aires.* nicolasraus@hotmail.com.

Facundo Luján. *GEMIS. Universidad Tecnológica Nacional, Regional Buenos Aires.* lujan.facundo@hotmail.com.

Cinthia Vegega. *GEMIS. Universidad Tecnológica Nacional, Regional Buenos Aires.* cinthiavg@yahoo.com.ar.

Pablo Pytel. *GEMIS. Universidad Tecnológica Nacional, Regional Buenos Aires. - GISI. Universidad Nacional de Lanús.* ppytel@gmail.com.

María Florencia Pollo-Cattaneo. *GEMIS. Universidad Tecnológica Nacional, Regional Buenos Aires.* fpollo@posgrado.frba.utn.edu.ar.