



UNIVERSIDAD NACIONAL DE LA MATANZA

ESCUELA DE POSGRADO

MAESTRIA EN INFORMÁTICA

TESIS DE MAESTRÍA

***“Extracción de conocimiento de foros
de aprendizaje”***

Autora: Prof. Nirva Ana Carestia

Director: Mg. Graciela Font

Buenos Aires, Noviembre 2014

“El conocimiento sólo existe en las mentes que lo han comprendido y justificado mediante la reflexión” (Paul, 1990)

AGRADECIMIENTOS

INDICE

CAPÍTULO 1: PRESENTACIÓN Y CONTEXTO DEL ESTUDIO

1.1 INTRODUCCIÓN	15
1.2 PROBLEMA	17
1.3 APROXIMACIÓN A UNA SOLUCIÓN	18
1.4 AMBITO DE LA INVESTIGACIÓN	19
1.5 OBJETIVOS	20
1.6 METODOLOGÍA	21
1.6.1 Estudio de caso	21
1.6.2 Descripción del caso	22
1.6.3 Diseño de la Investigación	23
1.7 ORGANIZACIÓN	24

CAPÍTULO 2: E-LEARNING

2.1 INTRODUCCIÓN	27
2.2 CONSTRUCTIVISMO COOPERATIVO	28
2.3 PENSAMIENTO CRÍTICO.....	29
2.4 APRENDIZAJE AUTOORIENTADO	30
2.5 COMUNIDAD DE INVESTIGACIÓN.....	30
2.6 PRESENCIA SOCIAL.....	32
2.7 PRESENCIA COGNITIVA.....	35
2.8 PRESENCIA DOCENTE.....	38
2.8.1 Diseño y organización.....	39
2.8.2 Facilitar el discurso.....	41
2.8.3 Enseñanza directa.....	42
2.9 VISIÓN HOLÍSTICA.....	44
2.10 CALIDAD EN ENTORNOS EDUCATIVOS VIRTUALES.....	45
2.10.1 Calidad y e-learning.....	45
2.10.2 El ciclo de la calidad.....	46

CAPÍTULO 3: MINERÍA DE DATOS

3.1 INTRODUCCIÓN	51
3.2 DEFINICIÓN DE MINERÍA DE DATOS	51
3.3 TIPOS DE MODELOS	53
3.4 EL PROCESO DE DESCUBRIMIENTO DE CONOCIMIENTO.....	54

3.5 REGLAS DE ASOCIACIÓN	55
3.6 ÁRBOLES DE DECISIÓN	58
3.6.1 Algoritmo ID3	59
3.6.2 Weka	68
3.6.3 Algoritmo J48	68
3.6.4 Evaluación de clasificadores.....	71
3.7 APLICACIONES.....	71
CAPÍTULO 4: EVALUACIÓN DE LA PRESENCIA COGNITIVA	
4.1 INTRODUCCIÓN	75
4.2 IDA FALLAS MONGE.....	76
4.3 ARROYO ÁVILA, REYES LÓPEZ Y GARCÍA BENCOMO	78
4.4 MAGDALENA CATALINA BUSSONE.....	81
4.5 MARÍA ESTELA RIPA.....	84
4.6 APORTES A LA INVESTIGACIÓN.....	89
CAPÍTULO 5: EXTRACCIÓN DE CONOCIMIENTO	
5.1 INTRODUCCIÓN	93
5.2 CURSO ENSEÑAR Y APRENDER EN EL AULA VIRTUAL	94
5.2.1 Presentación y Objetivos.....	94
5.2.2 Contenidos y metodología	95
5.3 CURSO PRÁCTICAS EDUCATIVAS EN ENTORNOS VIRTUALES	96
5.3.1 Presentación y Objetivos	96
5.3.2 Contenidos y metodología	97
5.4 ANÁLISIS DE LA PRESENCIA COGNITIVA.....	98
5.4.1 Justificación de la unidad de análisis.....	98
5.4.2 Heurísticas para la categorización.....	99
5.5 REGLAS DE ASOCIACIÓN.....	100
5.5.1 Definición.....	100
5.5.2 Reglas significativas 'soporte' y 'confianza'	105
5.5.3 Aplicaciones.....	110
5.6 ÁRBOLES DE DECISIÓN.....	110
5.6.1 Nivel de Presencia Cognitiva.....	110
5.6.2 Algoritmo J48.....	115
5.6.2.1 Algoritmo J48 con modo Use training set.....	117
5.6.2.2 Evaluación del clasificador.....	119

5.6.2.3 Algoritmo J48 con modo Supplied test.....	123
5.6.3 Reglas.....	125
5.7 MEJORA CONTINUA.....	127
CAPÍTULO 6: CONCLUSIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN	
6.1 SÍNTESIS: EL PROBLEMA Y LA SOLUCIÓN	131
6.2 APORTES AL LOGRO DEL OBJETIVO	132
6.3 CONCLUSIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN	134
CAPÍTULO 7: BIBLIOGRAFÍA	139
ANEXO A (WEKA)	
A.1: INTRODUCCIÓN	147
A.2: ARCHIVOS EN WEKA.....	147
A.3: FICHEROS PARA NIVEL DE PRESENCIA COGNITIVA.....	149
A.4: EJECUCIÓN DE WEKA.....	151
A.5: MODO EXPLORER.....	154
A.5.1: Preprocesado de datos.....	155
A.5.2: Clasificación.....	160
A.5.3: Búsqueda de asociaciones.....	170
ANEXO B (Instrumentos y heurística)	
B.1: DEFINICIÓN DE LA UNIDAD DE ANÁLISIS	175
B.2: PROCEDIMIENTO DE CODIFICACION.....	175
B.3: PREPARACIÓN DE LOS DATOS.....	176
B.4: INSTRUMENTO PARA CATEGORIZAR.....	176
B.5: PROCEDIMIENTO PARA CATEGORIZAR.....	181
B.6: INSTRUMENTOS PARA RECOLECCIÓN.....	181
B.6.1: Reglas de asociación.....	181
B.6.2: Árboles de Decisión.....	183
ANEXO C (Foro Grupo G5)	
C.1: ARCHIVO PREPROCESADO GRUPO G5.....	187
FIGURAS	
Figura 2.1: Comunidad de aprendizaje	32

Figura 2.2: Modelo de Investigación Práctica.....	36
Figura 3.1: Raíz del árbol de decisión	63
Figura 3.2: Subárbol correspondiente a temperatura agradable	66
Figura 3.3: Árbol completo correspondiente a temperatura	67
Figura 3.4: Árboles iniciales para temperatura y humedad.....	69
Figura 3.5: Árbol final para humedad	70
Figura 4.1: Categoría hecho desencadenante, dentro del componente presencia cognitiva....	79
Figura 4.2: Categoría exploración, dentro del componente presencia cognitiva	80
Figura 4.3: Categoría <i>integración</i> , dentro del componente <i>presencia cognitiva</i>	80
Figura 4.4: Categoría <i>resolución</i> , dentro del componente <i>presencia cognitiva</i>	81
Figura 4.5: Ejemplo 1.	82
Figura 4.6: Ejemplo 2.	83
Figura 5.1: Curso de Postgrado Enseñar y Aprender en el Aula Virtual.....	95
Figura 5.2: Curso Prácticas Educativas en Entornos Virtuales.....	97
Figura 5.3: Presentación de la actividad grupal de la Unidad I- 2008.....	100
Figura 5.4: Presentación práctica de aprendizaje grupal de la unidad I - 2011.....	113
Figura 5.5: Consignas práctica de aprendizaje grupal de la unidad I.....	113
Figura 5.6: Selección del algoritmo J48.....	118
Figura 5.7: Selección del modo de evaluación (Test Options).....	118
Figura 5.8: Información de ejecución.	119
Figura 5.9: Modelo de clasificación.....	119
Figura 5.10: Resumen	120
Figura 5.11: Precisión detallada por clase	121
Figura 5.12: Matriz de confusión.....	121
Figura 5.13: Lista de resultados.....	122
Figura 5.14: Árbol de clasificación.....	122
Figura 5.15: Selección del modo Supplied test set.....	123
Figura 5.16: Información de ejecución.....	124
Figura 5.17 Modelo de clasificación.....	124
Figura 5.18: Resumen.....	125
Figura 5.19 Precisión detallada por clase.....	125
Figura 5.20 Matriz de confusión.....	125
Figura A.1: Ventana de selección de Interfaz.....	152
Figura A.2 Ventana principal Simple CLI.....	152
Figura A.3: Ventana principal Explorer.....	153
Figura A.4: Ventana principal Experimenter.....	153

Figura A.5: Ventana principal Knowledge flow.....	154
Figura A.6 Ventana principal de preprocesado.....	157
Figura A.7: Ventana de Visualize all.....	159
Figura A.8: Ventana principal de Classify.....	162
Figura A.9: Ventana que permite la selección de clasificadores.....	163
Figura A.10: Selección del Modo de evaluación.....	164
Figura A.11 Opciones adicionales.....	164
Figura A.12 Información de ejecución.....	165
Figura A.13 Modelo de clasificación.....	165
Figura A.14 Resumen.....	166
Figura A.15 Precisión detallada por clase.....	167
Figura A.16 Matriz de confusión.....	168
Figura A.17 Lista de resultados.....	168
Figura A.18 Árbol de clasificación.....	169
Figura A.19 Ventana principal de Associate.....	170
Figura A.20 Parámetros del algoritmo Apriori.....	171
Figura A.21 Reglas generadas con el algoritmo A priori.....	171

TABLAS

Tabla 2.1: Indicadores, definiciones y ejemplos de las categorías de Presencia Social	34
Tabla 2.2: Descriptores e indicadores de las fases de investigación práctica.....	38
Tabla 2.3: Indicadores y ejemplos de diseño y organización.....	40
Tabla 2.4: Indicadores y ejemplos de facilitar el discurso.....	42
Tabla 2.5: Indicadores y ejemplos de enseñanza directa	43
Tabla 3.1: Base de datos con 4 productos y 5 transacciones	55
Tabla 3.2: Resumen de condiciones meteorológicas para las mañanas del día domingo.....	60
Tabla 3.3: Resumen para temperatura agradable	63
Tabla 3.4: Resumen para temperatura agradable y humedad normal	65
Tabla 4.1: Fases del modelo de investigación práctica	77
Tabla 4.2: Clasificación de los mensajes.....	77
Tabla 4.3: Síntesis Presencias.....	84
Tabla 4.4: Frecuencia y porcentajes de mensajes por categoría.....	86
Tabla 5.1: Indicadores de Presencia Cognitiva para los mensajes de la primera consigna	102
Tabla 5.2: Indicadores de Presencia Cognitiva para los mensajes de la segunda consigna	103
Tabla 5.3: Indicadores de Presencia Cognitiva para los mensajes de la tercera consigna.....	103
Tabla 5.4: Indicadores de Presencia Cognitiva para los mensajes de la cuarta consigna	103

Tabla 5.5: Indicadores de Presencia Cognitiva para los mensajes de la quinta consigna.....	103
Tabla 5.6: Indicadores de Presencia Cognitiva para los mensajes de la sexta consigna	104
Tabla 5.7: Indicadores de Presencia Cognitiva Final Grupo 1.....	104
Tabla 5.8: Indicadores de Presencia Cognitiva Final 3 Grupos.....	104
Tabla 5.9: Mensajes con Hb.....	106
Tabla 5.10: Mensajes del conjunto ($\{Hb\} \cup \{Eb\}$).....	108
Tabla 5.11: Mensajes con Eb.....	108
Tabla 5.12: Mensajes del conjunto ($\{Eb\} \cup \{Ia\}$).....	109
Tabla 5.13: Conjunto de entrenamiento para la clasificación	112
Tabla 5.14: Indicadores de Presencia Cognitiva Grupo 5.....	113
Tabla 5.15: Indicadores de Presencia Cognitiva Estudiante E1.....	114
Tabla 5.16: Indicadores y Nivel de Presencia Cognitiva Grupos 5 y 3.....	115
Tabla B.1: Indicadores de Presencia Cognitiva.....	176
Tabla B.2: Instrumento para recolección de indicadores por mensaje.....	182
Tabla B.3: Instrumento para recolección de datos estudiante E1.....	183
Tabla B.4: Instrumento para recolección de datos estudiante E2.....	184
Tabla B.5: Instrumento para recolección de datos estudiante E3.....	184
Tabla B.6: Instrumento para recolección de datos estudiante E4.....	184
Tabla B.7: Instrumento para evaluación nivel de presencia cognitiva	185

CAPÍTULO 1

Presentación y contexto del estudio

CAPÍTULO 1: PRESENTACIÓN Y CONTEXTO DEL ESTUDIO

1.1. INTRODUCCIÓN

El avance de las tecnologías de la información y la comunicación propician el establecimiento de comunidades de aprendizaje en las organizaciones educativas de nivel superior. Además, brindan la posibilidad de analizar nuevos escenarios para las prácticas docentes. Estas herramientas tecnológicas representan avances en los entornos educativos que permiten satisfacer los requerimientos de la sociedad del conocimiento a través de propuestas creativas e innovadoras.

Los nuevos entornos virtuales educativos exigen la introducción de cambios que requieren la revisión de las teorías y prácticas pedagógicas tradicionales. Los autores canadienses Garrison y Anderson (2005) ofrecen un marco teórico para la aplicación de la educación en línea o *e-learning*, entendido en el sentido amplio de “educación facilitada en línea mediante tecnologías en red, lo que no excluye el uso de otras tecnologías ni enfoques, incluidos las experiencias educativas presenciales” (p. 11). Ellos afirman que “el e-learning puede crear comunidades de investigación simultáneas accesibles para los estudiantes *en cualquier momento y en cualquier lugar* y que esta tecnología combinada con la pedagogía adecuada y un plan docente meditado, transformará la educación superior” (Garrison y Anderson, p. 14).

Algunos han calificado de explosivo el crecimiento del e-learning. Es cierto que se trata de un hecho sin precedentes, sorprendente e inaudito. De hecho, hay quien habla ya de una revolución en la educación superior. Otros sugieren que la tecnología del e-learning es única y que representa una nueva era en la educación a distancia (Garrison y Anderson, 2005).

“Teniendo en cuenta la adopción masiva del e-learning, lo que es sorprendente y motivo de preocupación es que se conozca tan poco sobre el uso de este medio” (Gilbert, 2000 citado por Garrison y Anderson, 2005, p. 12). Se trata entonces de encontrar métodos que puedan conducir a una comprensión del e-learning aplicados al ámbito de la educación superior.

Los ambientes virtuales de aprendizaje almacenan una gran cantidad de datos sobre las actividades de los estudiantes cuando estos toman un curso y usualmente esta información es utili-

zada para monitorear características del curso. La información se presenta comúnmente en formato tabular que, dependiendo de la cantidad o del tipo de datos, a veces resulta difícil de interpretar.

La información reduce la incertidumbre sobre algún aspecto de la realidad y, por tanto, permite tomar mejores decisiones. Prácticamente, no existe hoy en día una faceta de la realidad de la cual no se disponga de la información de manera electrónica, ya sea estructurada, en forma de base de datos, o no estructurada, en forma textual o hipertextual. Desgraciadamente, gran parte de esta información se genera con un fin concreto y posteriormente no se analiza ni integra con el resto de información o conocimiento del dominio de actuación.

Las técnicas de minería de datos o *data mining* pueden ayudar a la comprensión del e-learning y permitir a los docentes descubrir conocimiento para implementar estrategias de enseñanza-aprendizaje que permitan la mejora continua del proceso.

La minería de datos es una de las áreas de investigación que ha experimentado un crecimiento espectacular en los últimos años, ofreciendo herramientas potentes para el análisis de grandes bases de datos utilizadas en la empresas, industrias y ciencias (Zytkow y Klosgen, 2001).

Witten y Frank (2000) definen la minería de datos como “el proceso de extraer conocimiento útil y comprensible, previamente desconocido, desde grandes cantidades de datos almacenados en distintos formatos” (p. 5). Es decir, la tarea fundamental de la minería de datos es encontrar modelos inteligibles a partir de los datos. Para que este proceso sea efectivo debería ser automático o semiautomático y el uso de los patrones descubiertos deberá ayudar a tomar decisiones más seguras que reporten, por lo tanto, algún beneficio a la organización (Hernández, Ramírez y Ferri, 2004)

Data mining se aplica con éxito en sistemas de comercio electrónico, para comprender el comportamiento de los clientes y poder incrementar las ventas (Srivastava, Mobasher y Cooley, 2000). Las herramientas de data mining utilizan técnicas de extracción de conocimiento con el objetivo de descubrir información útil para la mejora de los sistemas. Aunque los métodos de descubrimiento de información aplicados en las áreas comerciales y

educativas son similares, los objetivos finales tienen matices diferentes debido a que en comercio electrónico el objetivo es guiar a los clientes durante la compra para maximizarla, mientras que en entornos virtuales educativos el objetivo es guiar a los estudiantes durante su aprendizaje para maximizarlo.

1.2. PROBLEMA

Según Dewey, “la interacción es el componente que define el proceso educativo y tiene lugar cuando los estudiantes transforman la información inerte que se les transmite, en conocimiento con valor y aplicaciones personales” (Dewey, 1916 citado por Garrison y Anderson, 2005, p. 65). “La interacción entre profesores y estudiantes, la interacción entre los propios estudiantes y la colaboración en el aprendizaje que resulta de esas interacciones son elementos clave en un proceso educativo” (Pallof y Prat, 1999 citado por Garrison y Anderson, p. 66).

Garrison y Anderson (2005), establecen que la interacción mediada tecnológicamente permite mantener el ritmo de aprendizaje, sirve para desarrollar relaciones entre los nuevos contenidos y los esquemas mentales existentes, permite configurar y reforzar la adquisición de nuevas capacidades, permite al profesor guiar el modo en que los estudiantes interactúan entre ellos y con los contenidos y facilita a los estudiantes la consecución de sus propios intereses.

Debido a la importancia central de la interacción en un entorno de educación no presencial, se considera relevante analizar el papel que desempeñan los foros en los entornos virtuales, ya que éstos proveen el principal espacio para el intercambio académico y sin embargo muchas veces no son planificados adecuadamente, ya que como mencionan Gros y Silva (2006), “en muchos casos la colaboración es vista desde una perspectiva superficial y se da por supuesto que el simple hecho de que un grupo de estudiantes intervengan en un foro virtual es sinónimo de aprendizaje y colaboración” (p. 2).

Estudios como el de Gunawardena y Mc. Isaac (2004), en los que se ha investigado la colaboración a partir de los datos cuantitativos de las intervenciones, no han permitido en la mayoría de los casos, más que tener una visión muy general de las cantidades y flujos de las interacciones, sin entrar en el contenido de la interacción y las consecuencias respecto al

aprendizaje. En este sentido, Perkins y Murphy (2006) sugieren “estudiar la contribución que las discusiones en línea puedan hacer en la promoción de estrategias de pensamiento crítico, como un ejemplo del tipo de investigaciones recomendables para tratar de entender el rol de estas tecnologías en la promoción del aprendizaje” (p. 298).

Dentro de diferentes modelos propuestos para el análisis de las participaciones en línea, Garrison y Anderson (2005) han estudiado la importancia del contexto y la creación de las comunidades de aprendizaje para facilitar la reflexión y el discurso crítico. En una comunidad centrada en la indagación, estos autores proponen la existencia de tres elementos que intervienen en un proceso de aprendizaje virtual y que son fundamentales para lograr el aprendizaje: la presencia cognitiva, la presencia social y la presencia docente. Para efectos del presente estudio se ha seleccionado la presencia cognitiva, que es una condición del pensamiento y el aprendizaje de alto nivel, y que hace referencia al “punto hasta el cual los estudiantes son capaces de construir significado, mediante la reflexión continua en una comunidad de investigación crítica” (Garrison, Anderson y Archer, 2001, p. 11).

1.3. APROXIMACIÓN A UNA SOLUCIÓN

En función del problema planteado, el desafío es extraer conocimiento de los foros de aprendizaje de una propuesta de e-learning que permita evaluar el nivel de presencia cognitiva, controlar el proceso de enseñanza-aprendizaje y apoyar la toma de decisiones.

Una de las aplicaciones más importantes de la minería de datos es el descubrimiento de reglas de asociación (Hipp, 2000) en el que se identifican relaciones entre elementos que suelen aparecer conjuntamente en un dominio determinado. Existen diferentes aproximaciones a este problema, dando lugar a distintos tipos de reglas de asociación: reglas de asociación cuantitativas, reglas de asociación negativa, reglas causales, reglas generalizadas o de predicción, etc.

El descubrimiento de reglas de asociación se ha planteado frecuentemente en sistemas de comercio electrónico, en los que se han buscado reglas del tipo: *Si un cliente compra el producto x_1 y x_2 , entonces comprará el producto x_3 con una determinada probabilidad.*

Debido a esta aplicabilidad y a su comprensión inherente, el establecimiento de reglas de asociación se ha convertido en un método muy popular de minería de datos.

Otro de los métodos de aprendizaje utilizados en minería de datos son los sistemas de aprendizaje basados en árboles de decisión. “Un árbol de decisión es un conjunto de condiciones organizadas en una estructura jerárquica, de tal manera que la decisión final a tomar se puede determinar siguiendo las condiciones que se cumplen desde la raíz del árbol hasta alguna de sus hojas” (Hernández et al., 2004, p. 282).

Estos árboles son especialmente apropiados para los procesos de decisión involucrados en tareas inductivas de clasificación. Clasificar es determinar de entre varias clases a que clase pertenece un objeto, la estructura de condición y ramificación de un árbol de decisión es idónea para este problema (Hernández et al., 2004). Se aplica en diversas áreas como: diagnóstico médico, juegos, predicción meteorológica, análisis de riesgos en la concesión de créditos, control de calidad, entre otros.

En esta tesis se propone:

- Analizar la presencia cognitiva en los foros de dos cursos de e-learning.
- Aplicar dos técnicas de minería de datos para extraer conocimiento de los foros de aprendizaje.
- Identificar las categorías que brindan mayor información respecto al nivel de presencia cognitiva.

1.4. AMBITO DE LA INVESTIGACIÓN

El estudio se enfoca a dos áreas principales: *minería de datos y aprendizaje*. En ésta última se circunscribe a las particularidades del e-learning.

En el ámbito de la minería de datos se utilizarán dos técnicas específicas: reglas de asociación y árboles de decisión.

Las particularidades del e-learning y los datos a procesar con las técnicas de minería de datos se obtienen de los siguientes proyectos de investigación:

Proyecto: *Categorizando la presencia cognitiva en la No-Presencialidad* Código F 841

Programa: *Desarrollo y Aplicación de Nuevas Tecnologías Educativo – Informáticas en Educación* – Res. N° 67/96-CD-FFHA Vigencia: 2008-2010 Aprobado por Res. N° 34/08-CS - Prórroga Res. N° 52/09-CS Directora: Mgter. Waisman, Elena Rut.

Proyecto: *Vicisitudes de la interacción: la enseñanza en entornos virtuales* Código F879

Programa: *Desarrollo y Aplicación de Nuevas Tecnologías Educativo – Informáticas en Educación* – Res. N° 67/96-CD-FFHA Vigencia: 01/01/2011 – 31/12/2013 Aprobado por Res. 037/11-CS – Prórroga Res. 131/12-CS. Directora: Mgter. Elena R. Waisman –

Limitaciones:

- El estudio abarca los foros de aprendizaje grupales de dos cursos de e-learning. El primero denominado *Enseñar y Aprender en el Aula Virtual* (2008) y el segundo *Prácticas Educativas en Entornos Virtuales* (2011).
- La investigación se limita al descubrimiento de reglas de asociación entre los indicadores de presencia cognitiva definidos por Garrison y Anderson (2005) en las interacciones de los foros de aprendizaje de tres grupos de estudiantes y a la construcción de un árbol de decisión que defina el nivel de presencia cognitiva alcanzado por los mismos. No se pretende evaluar los cursos en sí mismos.

1.5. OBJETIVOS

Objetivo general:

- Extraer conocimiento de los foros de dos cursos de e-learning con el objeto de apoyar la mejora continua del proceso de enseñanza-aprendizaje.

Objetivos específicos:

- Descubrir reglas de asociación entre los indicadores de presencia cognitiva.
- Construir un árbol de decisión que permita evaluar el nivel de presencia cognitiva en los foros de aprendizaje.

- Identificar las categorías que brindan mayor información respecto al nivel de presencia cognitiva.

1.6. METODOLOGÍA

En los siguientes incisos se describen la metodología y el diseño de la investigación utilizado en esta tesis.

1.6.1. Estudio de caso

La investigación se aborda desde la perspectiva de un *Estudio de Caso Mixto, Evolutivo y Colectivo* según se describe a continuación.

El estudio de caso se podría definir como “Estudios que al utilizar los procesos de investigación cuantitativo, cualitativo o mixto; analizan profundamente una unidad para responder al planteamiento del problema, probar hipótesis y desarrollar alguna teoría” (Hernández, Collado y Baptista, 2008, p. 2).

Por su parte Mertens (2005, citado por Hernández et al., 2008) define al estudio de caso “como una investigación sobre un individuo, grupo, organización, comunidad o sociedad; que es visto y analizado como una entidad” (p. 2).

La U. S. General Accounting Office, en 1990, proporciona una definición de estudio de caso: “constituye un método para aprender respecto a una instancia compleja, basado en un entendimiento comprehensivo de esta instancia como un todo y su contexto, mediante datos e información obtenidos por descripciones y análisis extensivos” (Mertens, 2005 citado por Hernández et al., 2008, p. 2).

Para Wiersma y Jurs (2005, citados por Hernández et al., 2008) el estudio de caso es “el examen detallado de *algo*: un evento específico, una organización, un sistema educativo, por ejemplo” (p. 2). En términos de Williams, Unrau y Grinnell (2005), el estudio de caso se concentra en una unidad de análisis. Yin (2003, citado por Hernández et al., 2008) señala que “un estudio de caso es una indagación empírica que investiga un fenómeno contemporáneo dentro de su contexto en la vida real, especialmente cuando los límites entre el fenómeno y el contexto no son claramente evidentes” (p. 2).

Stake (2000) utiliza el criterio de que el estudio de caso no está definido por un método específico, sino por su objeto de estudio. Entre más concreto y único sea éste, y constituya un sistema propio, con mayor razón se puede denominar estudio de caso (Citado por Hernández et al., 2008).

Varios autores como Stake (2003), Mertens (2005), Williams, Grinnell y Unrau (2005) opinan que más que un método es un diseño y una muestra, argumentan que los estudios de caso utilizan o pueden utilizar diversos métodos.

Hernández et al. (2008) dicen que “su importancia -más que discutir sobre si es un método, un diseño o una muestra- reside en su utilización”. Además indican que:

Cada vez son más comunes los estudios de caso mixtos, de hecho, la sugerencia que podemos hacer es que resulta conveniente que sean híbridos, recolecten tanto datos cuantitativos como cualitativos; lo cual fortalece su amplitud y profundidad. En este sentido, un estudio de caso puede ser en dos etapas cuantitativa-cualitativa o viceversa, de enfoque principal, en paralelo o mixto complejo. Asimismo, aplican los criterios cuantitativos y cualitativos para establecer la confiabilidad y validez de la investigación. (p. 22)

Por su finalidad Stake (2000, citado por Hernández et al. 2008) identifica tres diferentes tipos de estudios de caso: *intrínsecos*, *instrumentales* y *colectivos*.

El propósito de los *primeros* no es construir una teoría, sino que el caso mismo resulte de interés. Los estudios de casos *instrumentales* se examinan para proveer de insumos de conocimiento a algún tema o problema de investigación, refinar una teoría o aprender a trabajar con otros casos similares. Por su parte, los *colectivos* sirven para construir un cuerpo teórico,-sumar hallazgos, encontrar elementos comunes y diferencias, así como acumular información. (p. 6)

Por su temporalidad pueden ser: a) temporales: duración no prolongada, regularmente un año o menos y b) longitudinales o evolutivos: más de un año y varias etapas de recolección de los datos o mediciones.

1.6.2. Descripción del caso

El estudio abarca los foros de aprendizaje de dos cursos de e-learning similares, dirigidos a docentes universitarios de diferentes disciplinas, implementados en la plataforma *Moodle* y organizados por el mismo grupo de profesionales. La palabra Moodle es un acrónimo de

Module Object-Oriented Dynamic Learning Environment cuya traducción es Entorno Modular de Aprendizaje Dinámico Orientado a Objetivos.

El primero de los cursos denominado *Enseñar y Aprender en el Aula Virtual* se desarrolla durante el periodo agosto-diciembre de 2008 y el segundo *Prácticas Educativas en Entornos Virtuales* en el periodo abril-julio de 2011.

En esta investigación se propone la aplicación de dos técnicas de minería de datos para el análisis de la Presencia Cognitiva en los foros de aprendizaje de tres grupos de estudiantes de los dos cursos de e-learning.

1.6.3. Diseño de la investigación.

El término *diseño* se refiere al “esbozo, prototipo o modelo que indica el conjunto de decisiones, pasos y actividades que se realizarán para guiar el curso de una investigación” (Ander-Egg, 2000, p. 90). De allí la importancia de definir el diseño de la investigación, pues implica trazar la estrategia para el logro de los objetivos de la investigación y para responder a los interrogantes planteados.

Se planifican las siguientes actividades:

- Seleccionar bibliografía especializada sobre la temática.
- Indagar las características del e-learning en el ámbito de la educación superior.
- Describir dos técnicas de minería de datos: reglas de asociación y árboles de decisión.
- Analizar experiencias de evaluación de Presencia Cognitiva.
- Compilar y organizar los mensajes de los foros.
- Seleccionar e identificar los mensajes a categorizar.
- Definir las heurísticas que se utilizan para analizar los mensajes de los foros.
- Diseñar los instrumentos para la recopilación de datos.
- Diseñar el instrumento a aplicar para la categorización de Presencia Cognitiva.
- Aplicar la técnica de reglas de asociación.
- Aplicar la técnica de árboles de decisión.
- Utilizar el software WEKA (Waikato Environment for Knowledge Analysis) para probar los algoritmos de clasificación.

- Elaborar Conclusiones.

1.7. ORGANIZACIÓN

La tesis se organiza en siete capítulos y tres anexos. A continuación se da una descripción general de cada uno.

Capítulo 1: Introduce en la temática, plantea el problema, la posible solución y los objetivos de la investigación. Además, describe la metodología.

Capítulo 2: Presenta una descripción del e-learning y las características del modelo de comunidad de investigación en la educación superior.

Capítulo 3: Define minería de datos y describe las técnica de reglas de asociación y árboles de decisión.

Capítulo 4: Describe experiencias de evaluación de la Presencia Cognitiva en las que se aplica el mismo modelo de análisis.

Capítulo 5: Describe el *caso* que se estudia y el procedimiento seguido para analizarlo con cada una de las técnicas aplicadas. Presenta los resultados y sus posibles aplicaciones.

Capítulo 6: Incluye una breve descripción del problema y de la solución propuesta, esboza los beneficios y limitaciones de la misma, presenta las conclusiones finales y propone trabajos de investigación futuros.

Capítulo 7: Detalla la bibliografía citada en la tesis.

Anexos: Incluye material creado por la tesista.

CAPÍTULO 2

E-Learning

CAPÍTULO 2: E-LEARNING

2.1. INTRODUCCIÓN

La incorporación de las tecnologías de información y comunicación en el ámbito educativo han provocado la aparición de diferentes conceptos que tratan de caracterizar estos *nuevos escenarios de enseñanza-aprendizaje*. En este campo de estudio, se utilizan con escasas diferencias y a veces como sinónimos, una diversidad de términos. Entre ellos, se han identificado:

entornos virtuales de enseñanza-aprendizaje (EVE-A); entornos virtuales de aprendizaje (EVA); entornos hipermedia o entornos telemáticos; ambientes virtuales de aprendizaje (AVAs); aprendizaje con sistemas inteligentes; páginas webs educativas; enseñanza asistida por ordenador (EAO); nuevos entornos de aprendizajes con TIC; ambientes de aprendizajes colaborativos; ambiente virtual colaborativo (AVC); e-learning; educación en campus virtuales; aprendizaje en red; formación en red; comunicación mediada por Ordenador (CMO). (Gewerc, 2008, pp. 18 -19)

El término *e-learning* es uno de los más utilizados, según Gewerc (2008),

la ‘e’ responde al amplio espectro de recursos que las redes electrónicas ponen a nuestro alcance y en las que el aprendizaje puede tener lugar. Se refiere al hardware ordenadores de mesa, ordenadores portátiles, teléfonos móviles, reproductores de mp3, etc. y también al software que pueda codificar, recolectar, almacenar, enviar y presentar información/comunicación en forma de textos, imágenes, audio; aplicaciones y sistemas que brindan la oportunidad de compartir herramientas para apoyar la información, gestionar los ambientes de los cursos, comunicar mediante el ordenador, y trabajar de forma colaborativa en ambientes virtuales. La palabra *learning* se refiere a una metodología de construcción de conocimientos y desarrollo de habilidades centradas en el sujeto que aprende, y no tanto en el profesor que enseña *teaching*. (p. 19)

Sin embargo, y a pesar de utilizar el término *e-learning* por ser uno de los más consensuados, no implica desdibujar el rol de la enseñanza ni el de los docentes. Por esa razón, una definición en profundidad del término conduce, a precisiones sobre lo que se entiende por aprendizaje en el contexto tecnológico, ya que el alcance y aplicación del término dependen de la o las teorías que se utilicen.

Los autores canadienses Garrison y Anderson (2005) ofrecen un marco para la aplicación de e-learning en el ámbito de la educación superior y definen el e-learning como “la educación facilitada on line mediante tecnologías en red” (p. 11).

Los mismos autores destacan que el potencial del e-learning está en la interacción y la comunicación y no se limita al hecho de facilitar el acceso a la información. El objetivo de un e-learning de calidad es la formación dinámica e intelectualmente estimulante. La interacción supera la transmisión unilateral de contenidos y apoya la comunicación entre los actores del proceso educativo.

Además plantean que el desafío real está en entender la naturaleza y el potencial del e-learning y sus implicaciones para un nuevo modelo de enseñanza. La educación superior está tratando de superar el individualismo, el aislamiento y la transmisión de contenidos, orientándose a los enfoques asociados con el aprendizaje en comunidad y la construcción personal del significado. La tecnología del e-learning permite la aplicación de los nuevos modelos y el mejoramiento del proceso de enseñanza-aprendizaje.

El presente trabajo se sustenta en el marco teórico propuesto por Garrison y Anderson (2005). En el resto del capítulo se esbozan los conceptos que permiten comprender y aplicar el e-learning con fines educativos desde la perspectiva adoptada.

2.2. CONSTRUCTIVISMO COOPERATIVO

El constructivismo cooperativo plantea que existe una relación entre la construcción personal de significado y la influencia que ejerce la sociedad en la configuración de la relación educativa (Garrison y Archer, 2000). Este proceso reconoce la interacción entre el significado individual y el conocimiento socialmente construido. La identificación de estos dos intereses es fundamental para la construcción de un marco teórico mediante el cual se puede comprender y aplicar el e-learning con fines educativos.

Desde el punto de vista filosófico, esta perspectiva constructivista cooperativa está asociada con el trabajo de John Dewey. Este autor rechazaba todo pensamiento dualista, sobre todo en relación con el individuo y la sociedad. Para Dewey, (citado por Garrison y Anderson, 2005) “la sociedad y el individuo no pueden existir ni por separado ni en relación de subordinación” (p. 31). Comprender que existe una interacción entre la experiencia y los intereses personales, por un lado, y los valores, normas y conocimientos sociales, por otro, es lo que permite entender la perspectiva constructivista cooperativa de la educación.

Dewey (1938 citado por Garrison y Anderson, 2005) identificó dos principios el de *interacción* y el de *continuidad*.

El primero unifica en un marco temporal inmediato los mundos subjetivo (personal) y objetivo (social). Mediante esta interacción se crean las ideas que dan sentido al mundo exterior. Es decir, el significado es construido y compartido. A través de la interacción, las ideas son comunicadas y el conocimiento construido y confirmado. La continuidad es la capacidad del individuo de seguir formándose. La importancia del segundo principio es que puede constituirse como la base de la educación futura y tiene un gran valor tanto para el individuo como para la sociedad. (p. 31)

La perspectiva filosófica del constructivismo cooperativo define la transacción educativa. Es decir, la cooperación y el constructivismo corresponden a las responsabilidades de profesores y estudiantes en una experiencia de formación. La interacción profesor/alumno es una representación adecuada de una experiencia educativa cooperativa y constructivista (Garrison y Anderson, 2005).

2.3. PENSAMIENTO CRÍTICO

El pensamiento crítico es un modelo cognitivo que se proyecta desde el interior hacia el exterior. Este modelo refleja las distintas fases que se reproducen entre las esferas privada y pública del individuo. Las fases del pensamiento crítico son el hecho desencadenante, la exploración, la integración y la resolución (Garrison y Anderson, 2005)

El concepto de pensamiento crítico deriva del modelo de pensamiento reflexivo de Dewey (1933 citado por Garrison y Anderson, 2005). Según este autor,

el pensamiento reflexivo o crítico tiene valor práctico en la medida en que profundiza en el significado de las experiencias personales representando así un objetivo educativo central. El pensamiento crítico autentifica el conocimiento existente y al tiempo genera nuevo conocimiento, lo cual sugiere una conexión íntima con la educación. (p. 85)

El pensamiento crítico es sinónimo de voluntad de indagación, de investigación. El concepto de pensamiento reflexivo de Dewey es comprensivo y coherente ya que la mayoría de las formas de pensamiento como el creativo, el crítico, el intuitivo pueden ser interpretadas dentro de ese marco (Garrison y Archer, 2000).

2.4. APRENDIZAJE AUTOORIENTADO

El aprendizaje autoorientado tiene relación con las responsabilidades y estrategias de la gestión docente. Esta gestión incluye tanto aspectos técnicos como cognitivos. La organización de la información y de los recursos tecnológicos forman parte de la gestión técnica. El diseño, la organización y el seguimiento de las tareas intelectuales corresponden a la gestión cognitiva. Se trata de crear un entorno educativo que desarrolle y promueva la capacidad para pensar y aprender de forma autónoma y al mismo tiempo conjunta. El objetivo es superar la etapa de asimilación de información y lograr la formación de estudiantes críticos, autoorientados, motivados, que sean capaces de reflexionar, cooperar y seguir aprendiendo a lo largo de la vida (Garrison y Anderson, 2005).

2.5. COMUNIDAD DE INVESTIGACIÓN

El nuevo escenario para la educación superior está formado por comunidades de aprendizaje con sentido de cooperación. Los profesores y estudiantes que interactúan con el objetivo de facilitar, construir y validar el conocimiento, y de desarrollar capacidades que permitan continuar la formación en el futuro, constituyen una comunidad de aprendizaje.

Estas comunidades fomentan la independencia cognitiva y la interdependencia social, es decir la fusión del mundo individual o subjetivo y del mundo compartido u objetivo. La construcción del conocimiento individual se debe, en gran medida, al entorno social. Es decir, un entorno que ofrezca opciones y una diversidad de perspectivas puede promover la investigación crítica y creativa. Este tipo de comunidades son el elemento central del marco conceptual del e-learning y en un requisito para la educación de alto nivel (Garrison y Anderson, 2005).

Para Lipman (1991 citado por Garrison y Anderson 2005, p. 48) es necesaria la presencia de una comunidad de investigación para poner en práctica la metodología educativa del pensamiento crítico. Esta comunidad es guiada por profesores, pero no en forma autoritaria sino con el objetivo de promover la participación, ordenar las relaciones y moderar todo el proceso. La comunidad es el entorno fundamental para la investigación crítica personal y la construcción del significado. La colaboración del grupo y la responsabilidad individual dan

sentido a la experiencia educativa.

En una comunidad de investigación existe tanto racionalidad como libertad. Como afirma Lipman (1991 citado por Garrison y Anderson, 2005), una comunidad de investigación es donde:

Los estudiantes se escuchan unos a otros con respeto, se enriquecen mutuamente con las ideas que aportan, se exigen argumentos a favor de las opiniones que unos defienden y otros atacan, se ayudan para extraer conclusiones de lo que se dice e intentan identificar las asunciones de los demás. (p. 48)

Un contexto de e-learning ofrece ventajas únicas para la creación de comunidades de investigación. La comunicación de base textual, la oportunidad de acceder a fuentes inagotables de información y la capacidad para la interacción ofrecen el entorno apropiado para implementar este tipo de comunidades.

Los tres elementos básicos de una comunidad de investigación que deben ser tenidos en cuenta al planificar y estructurar una experiencia de e-learning son: *la presencia cognitiva, la presencia social y la presencia docente*.

La presencia cognitiva es una condición del pensamiento y el aprendizaje de alto nivel, se refiere a los resultados educativos. Es decir, la presencia cognitiva “es el punto hasta el cual los estudiantes son capaces de construir significado mediante la reflexión continua en una comunidad de investigación” (Garrison, Anderson y Archer, 2001, p. 11).

La presencia social “es la capacidad de los participantes en una comunidad de investigación de proyectarse a sí mismos social y emocionalmente, como personas ‘reales’, es decir, su personalidad plena, mediante los medios de comunicación en uso” (Garrison, Anderson y Archer, 2001, p. 11).

La presencia docente es definida como “la acción de diseñar, facilitar y orientar los procesos cognitivo y social con el objetivo de obtener resultados educativos personalmente significativos y de valor docente” (Garrison y Anderson, 2005, p. 51).

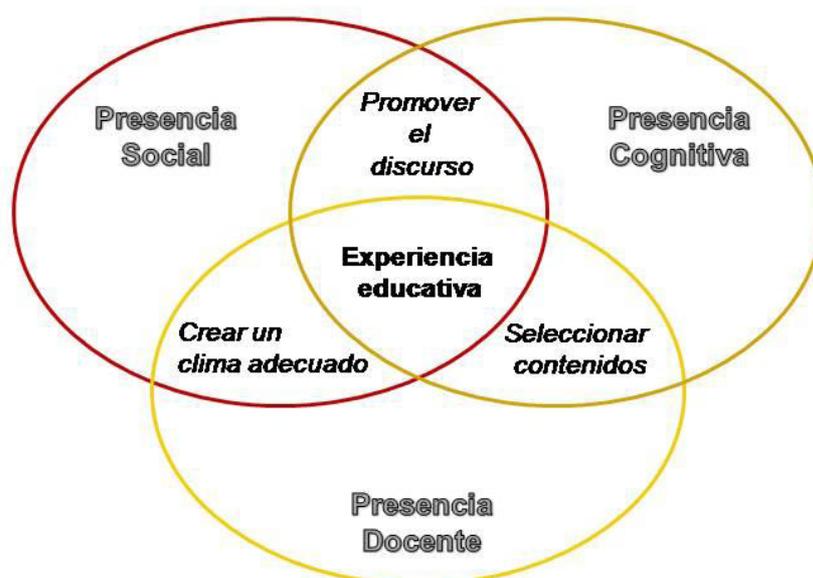


Figura 2.1. Comunidad de aprendizaje

Fuente: Arroyo, Reyes y García (2010, p. 7)

2.6. PRESENCIA SOCIAL

Los estudiantes están familiarizados con el uso social de los espacios virtuales. El desafío es utilizarlos para crear comunidades de investigación. Estas comunidades de estudiantes utilizarán un medio de comunicación asincrónico de base textual que representa un cambio cualitativo respecto a la comunicación en tiempo real, verbal y cara a cara de las clases presenciales (Garrison y Anderson, 2005).

El aislamiento, la soledad, la ausencia de lo comunitario se plantean como problemas latentes en educación a distancia. Se debe reflexionar sobre lo que significa la presencia social en una comunidad de e-learning caracterizada por su particular modo de comunicación. El aprendizaje no es una experiencia individual, consiste en la cooperación y ello implica un sentido de pertenencia y de aceptación en un grupo con intereses comunes.

Garrison y Anderson (2005) elaboran tres grandes categorías de indicadores de presencia social: *rasgos de comunicación afectiva*, *de comunicación abierta* y *de comunicación cohesiva* que se pueden ver con detalle en tabla 2.1. “Compartir la esfera socio-emocional es inherente a la funcionalidad comunicativa y a la cohesión de una comunidad de investigación” (p. 78).

La *comunicación afectiva* es una característica de la presencia social y de la participación en una comunidad de investigación. El interés y la constancia son factores fundamentales de toda experiencia educativa y son en cierta medida, muestras de afecto. Estas conductas facilitan el diálogo necesario para una experiencia formativa. Además, el respeto y el apoyo son necesarios para la reflexión y el discurso críticos. “Las muestras de afecto implican el reconocimiento tácito de la existencia de una relación recíproca con la comunidad” (Garrison y Anderson, 2005, p. 78).

La comunicación afectiva se puede detectar a través de tres indicadores. En primer lugar, la emoción puede expresarse mediante efectos visuales en el texto: letras mayúsculas, signos de puntuación, emoticones, etc. En segundo lugar, mediante la utilización del lenguaje para expresar estados de ánimo, a través de bromas, ironías, sarcasmo, etc. En tercer lugar, otra forma muy humana de establecer vínculos o relaciones de afecto es mediante la apertura al otro. En general, cuanto más se sabe sobre los otros miembros de la comunidad, más confianza y responsabilidad se siente. Se pueden ver ejemplos en tabla 2.1.

Garrison y Anderson (2005) explican que “la *comunicación abierta* consiste en generar respuestas pertinentes y constructivas a las cuestiones planteadas por los demás” (p. 80). Además, indican que es recíproca, respetuosa y posee calidad afectiva que se manifiesta en un clima de confianza y aceptación. Este tipo de comunicación permite cuestionar ciertas cosas a la vez que protege la autoestima y la aceptación en la comunidad. La comunicación abierta se construye mediante un proceso de reconocimiento y apreciación de las aportaciones de los otros, promoviendo así la participación y la interacción. Las expresiones de acuerdo o el cuestionarse el contenido de los mensajes, revela un cierto nivel de compromiso en el proceso de reflexión y de discurso crítico. En un entorno del e-learning, la comunicación, inherentemente reflexiva, se construye íntegramente sobre la base de la comunicación abierta.

Todos los indicadores anteriores contribuyen directamente a la tercera categoría de la presencia social: *la cohesión del grupo*. La cohesión del grupo es fundamental para mantener el compromiso y los objetivos de una comunidad de investigación, especialmente en un grupo de e-learning separado por el tiempo y el espacio. La construcción del significado y la confirmación de la comprensión sólo pueden obtenerse en una comunidad cohesionada. La calidad de los resultados educativos sólo puede lograrse cuando los estudiantes se perciban

como parte de una comunidad de investigación. Una forma de promover la cohesión es dirigirse a las personas por su nombre. También contribuyen a fortalecer la cohesión grupal la utilización de pronombres inclusivos como *nosotros* y expresiones del tipo *Hola a todos*, *Cómo estamos* (Garrison y Anderson, 2005).

Tabla 2.1. *Indicadores, definiciones y ejemplos de las categorías de Presencia Social*

Categoría	Indicadores	Definición	Ejemplo
Afecto	Expresión de emociones	Expresiones convencionales o no convencionales de emoción, como por ejemplo puntuación repetitiva, proliferación de mayúsculas, emoticones, etc.	¡Es que no soporto que...!!!!, ¡TODO EL MUNDO FUERA!
	Recurrir al Humor	Bromas, ironías, medias palabras, sarcasmo	Este año, la cosecha de plátanos en Calgary tiene buena pinta ;
	Expresarse Abiertamente	Presenta detalles de la vida fuera de la clase, o expresa vulnerabilidad	Eso es exactamente lo que hacemos en el trabajo... No entiendo su Pregunta
Comunicación abierta	Seguir el hilo	Emplear el signo de respuesta del software, en vez de iniciar un nuevo hilo de comunicación	Dependiente del software por ejemplo: «Asunto: Re o sección de
	Citar los Mensajes de otros	Emplear los recursos del software para citar os mensajes completos de otros o cortar y pegar fragmentos de otros mensajes	por ejemplo, «Marta escribe: o texto precedido por el símbolo
	Referirse Explícitamente a los mensajes de otros	Hacer referencia a los contenidos de otros mensajes	En tu mensaje, comentabas la distinción que hace Moore entre.
	Hacer preguntas	Los estudiantes hacen preguntas a otros estudiantes o al	¿Alguien más ha tratado con WEBCT?

Cohesión		moderador	
	Expresar Aprecio	Felicitar a los demás, hacer cumplidos, por el contenido de sus mensajes	Me gusta mucho tu interpretación de la lectura
	Expresar acuerdo	Expresar acuerdo con otros o con el contenido de sus mensajes	yo estaba pensando lo mismo. Diste en el clavo
	Vocativos	Dirigirse o referirse a los participantes por sus nombres	Creo que Juan planteó una cuestión interesante. Juan, ¿tú que opinas?
	Dirigirse o referirse al grupo usando pronombres inclusivos	Dirigirse al grupo como nosotros, nuestro grupo	Nuestro libro de texto se refiere a..., Creo que nos hemos desviado un poco
	Elementos fácticos, saludos	Comunicación que desempeña una función social; saludos, despedidas	Hola a todos, Esto es todo por ahora, Aquí tenemos muy buen tiempo

Fuente: Garrison y Anderson (2005, p. 79)

2.7. PRESENCIA COGNITIVA

Garrison y Anderson (2005) adoptan como marco teórico para el estudio del e-learning el modelo bidimensional de investigación práctica propuesto por Dewey en 1933, que se muestra en la figura 2.2.

La investigación práctica se basa en la experiencia. La integración de los mundos público y privado del estudiante es un concepto central para la creación de la presencia cognitiva con fines educativos. El continuo existente entre acción y deliberación queda reflejado en la primera dimensión del modelo. La transición entre los mundos concreto y abstracto se refleja en la dimensión percepción-concepción. Esta segunda dimensión recoge el proceso cognitivo que asocia hechos o acontecimientos con ideas o conceptos. (p. 88)

Este modelo identifica cuatro fases: activación, exploración, integración y resolución, que se utilizan para la descripción de la presencia cognitiva en un contexto educativo en general y en

el e-learning en particular. No se trata de fases inmutables, son generalizaciones que, en la práctica, pueden ser alteradas para lograr la comprensión. “Una comprensión metacognitiva de todas las fases puede ser de gran valor tanto para el profesor como para el estudiante a la hora de valorar las tareas en curso y los progresos realizados” (Garrison y Anderson, 2005, p. 88).

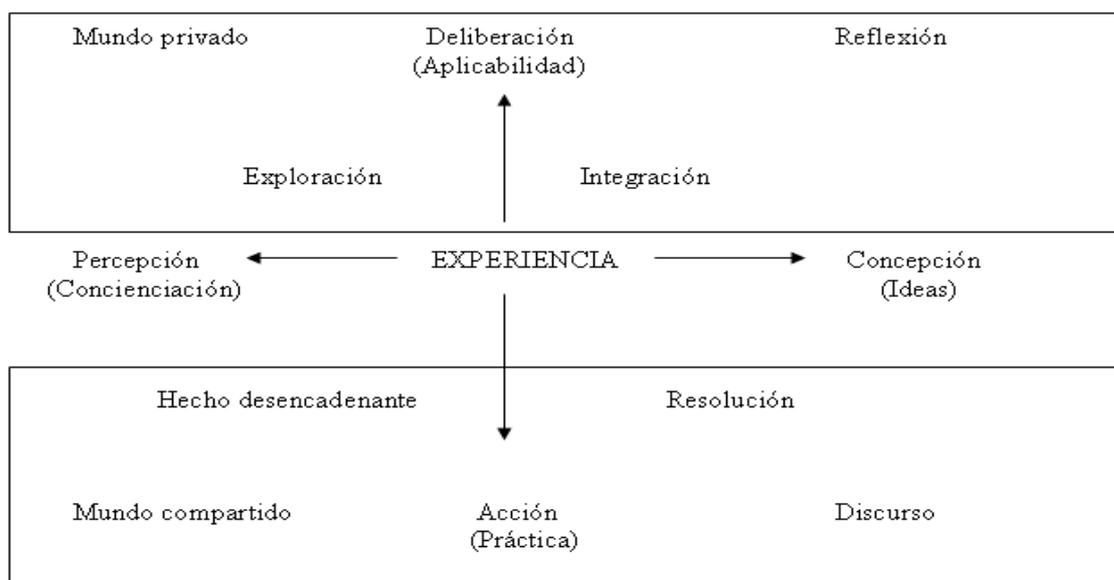


Figura 2.2. Modelo de Investigación Práctica

Fuente: Garrison y Anderson (2005, p. 89)

La primera fase de la investigación práctica radica en el hecho desencadenante. En una experiencia educativa esa fase debe consistir en alguna actividad pensada a conciencia para asegurar la plena implicación de los estudiantes. Debería, además, referirse a un asunto o concepto central. De ser posible, debería presentar un dilema o problema relacionado con la experiencia de los estudiantes o con sus estudios previos. Aunque es responsabilidad del profesor iniciar la fase de investigación, ese momento también puede estructurarse de un modo abierto. Podría, por ejemplo, presentar el tema y animar a los estudiantes a que planteen preguntas o problemas que ellos perciban o hayan vivido. Este planteamiento genera muchos resultados positivos en términos de la participación de los estudiantes, de la evaluación del estado del conocimiento y de la generación de ideas constructivas y no esperadas (Garrison y Anderson, 2005).

La segunda fase es la exploración. Consiste en entender la naturaleza del problema, buscar

información relevante y posibles explicaciones. Es un proceso inquisitivo y divergente que puede abordarse mediante actividades de grupo y/o mediante actividades más privadas como la investigación bibliográfica. Es posible que los estudiantes experimenten cierta redundancia entre el mundo reflexivo y el compartido en la medida en que las ideas son exploradas de modo cooperativo y los individuos intentan encontrar sentido a la aparente complejidad y confusión. Pero en eso consiste precisamente la esencia de una auténtica comunidad de investigación. Es necesario gestionar y orientar esta fase de pensamiento divergente para que sirva de transición para la fase siguiente (Garrison y Anderson, 2005).

La tercera fase corresponde a la integración y se orienta hacia la construcción del significado. Se trata de elaborar una solución o explicación apropiada mediante la integración y sistematización de ideas. Es una fase básicamente reflexiva, pero los estudiantes deben implicarse activamente en el discurso crítico para poder llegar a la comprensión. Las características particulares del e-learning, como su base textual y su naturaleza asincrónica, pueden facilitar la obtención de resultados significativos. Por estas razones, ésta es una fase especialmente decisiva para la creación de presencia cognitiva. El profesor debe evaluar el nivel de comprensión, detectar los posibles errores de concepto y orientar el proceso de pensamiento crítico. Generalmente se tiende al estancamiento en la fase de la exploración y a no avanzar hacia fases superiores. La creación de presencia cognitiva requiere la implicación de los estudiantes en todas las fases de la investigación práctica, incluyendo una apreciación metacognitiva de las distintas etapas y retos que se experimentan (Garrison y Anderson, 2005).

La cuarta fase es la resolución del dilema o problema, puede encontrarse una solución específica o reducirse la complejidad mediante un marco de significado. En esta fase de confirmación o de prueba se evalúa la viabilidad de la solución propuesta mediante su aplicación directa o indirecta. La confirmación directa es más difícil y suele ser poco práctica en un entorno educativo. La búsqueda indirecta o mental de soluciones es una actividad educativa viable y de gran valor, además al igual que en la vida real, es muy poco frecuente dar con soluciones definitivas. Los resultados de la fase de resolución suelen plantear nuevas preguntas, activando nuevos ciclos de indagación y, por tanto, promoviendo el aprendizaje continuo (Garrison y Anderson, 2005).

La tabla 2.2 presenta los descriptores e indicadores que corresponden a cada fase del proceso de investigación práctica.

Tabla 2.2. *Descriptores e indicadores de las fases de investigación práctica.*

Fase	Descriptor	Indicador
Hecho desencadenante	Evocativo (inductivo)	Reconocer el problema
		Confusión
Exploración	Inquisitivo (divergente)	Divergencia
		Intercambio de información
		Lluvia de ideas
		Salto intuitivos
Integración	Tentativo (convergente)	Convergencia
		Síntesis
		Soluciones
Resolución	Comprometido (deductivo)	Aplicar
		Comprobar
		Defender

Fuente: Garrison y Anderson (2005, p. 92)

2.8. PRESENCIA DOCENTE

En una experiencia educativa, tanto el profesor como los estudiantes tienen responsabilidades en el proceso del aprendizaje. A la presencia docente le corresponde garantizar el equilibrio transaccional adecuado, gestionar y hacer el seguimiento de todo el proceso y evaluar los resultados obtenidos.

La transacción educativa es muy compleja tanto en el entorno presencial como en el de e-learning pues presenta múltiples dimensiones. Las responsabilidades y principios básicos son los mismos en cualquier contexto; entre ellas se puede incluir “ser un experto en la materia que se enseñe, saber diseñar programas docentes, ser un animador social y un profesor” (Garrison y Anderson, 2005, p. 97). Pero el marco del e-learning altera de forma significativa el modo de desarrollar todas esas funciones.

Los roles del educador pueden clasificarse en tres categorías principales: diseño y organización, facilitar el discurso y enseñanza directa. En coherencia con ello, se define la presencia docente como “el diseño, facilitación y orientación de los procesos cognitivo y social con el objetivo de obtener resultados educativos significativos desde el punto de vista personal y docente” (Garrison y Anderson, 2005, p. 98).

La presencia docente, es lo que el profesor hace para crear una comunidad de investigación que integre la presencia social y la cognitiva y para que se mantenga un entorno educativo dinámico. A continuación se presentan descripciones detalladas de cada uno de los elementos de la presencia docente con el fin de que puedan ser utilizadas para su identificación.

2.8.1. DISEÑO Y ORGANIZACIÓN

El diseño y la organización se relacionan con la macroestructura y con el proceso. No es de extrañar que el diseño y organización de un curso mediante e-learning sea, por lo menos al principio, más difícil que el diseño y organización de un curso similar en un contexto tradicional. Ello se debe, en primer lugar, a la tecnología y a la necesidad de los profesores de rediseñar enfoques que maximicen las capacidades del medio. En algunos casos, el rediseño puede suponer un gran esfuerzo. Aquellos que, en el pasado, se hayan basado exclusivamente en dar clases, sentirán mucho más necesario centrarse en el aspecto de diseño y organización docente al desplazarse a un entorno de e-learning (Garrison y Anderson, 2005).

El diseño del programa de estudios se hace más complicado por tener que aumentar y, al mismo tiempo, reducir los contenidos. Aumentar en el sentido de facilitar enlaces a otros sitios de Internet que contengan objetos relevantes de estudio o material suplementario. Reducir en el sentido de que, si debe darse un nivel considerable de interactividad, hay que dar un nuevo sentido a los clásicos apuntes de clase y reducir la cantidad de material presentado. Junto a esta ampliación y canalización de los materiales docentes hay que destacar la tarea fundamental de seleccionar las actividades tanto individuales como de grupo y considerar cuánto tiempo debe dedicarse a cada una de ellas. Ahí es donde se da la intersección entre una buena comprensión del medio y la transacción educativa auténtica. El esfuerzo de diseño antes de iniciar el curso reportará numerosos beneficios que se reflejarán en los resultados. Ello no debe ser un obstáculo, sin embargo, para las decisiones de organización que deban tomarse sobre la marcha. La tabla 2.3 presenta los indicadores del

diseño junto a algunos ejemplos.

El diseño y la organización estructuran cualquier experiencia educativa e implican responsabilidades y funciones similares. Garrison y Anderson (2005) indican que:

La diferencia consiste en que el diseño se refiere a las decisiones estructurales adoptadas antes de que comience el proceso, mientras que la organización se refiere a decisiones similares que son tomadas para adaptarse a los cambios durante la transacción educativa. Los comentarios sobre el aspecto organizativo reflejan la naturaleza flexible y no prescriptiva de cualquier experiencia educativa. El diseño no es más que una plantilla flexible, creada con la expectativa de que, inevitablemente, surgirán necesidades y asuntos concretos y que habrá que hacer modificaciones sobre la marcha. (p. 99)

La interactividad del e-learning otorga cada vez más importancia a las cuestiones organizativas. La indeterminación de la entrada y desarrollo del conocimiento de los estudiantes introducirá, inevitablemente, alguna incertidumbre en el proceso de diseño. Si el e-learning debe ser un proceso cooperativo, constructivista, entonces los estudiantes deberían tener alguna influencia en lo que se estudia y en cómo se enfocan los contenidos. Por tanto, el diseño no debería separarse de la forma de presentación. Debe asegurarse una cierta continuidad entre la fase de diseño y la de organización. Todo ello se puede conseguir de modo óptimo cuando el profesor puede diseñar y al mismo tiempo organizar la experiencia educativa de modo que introduzca una responsabilidad efectiva hacia el desarrollo de las necesidades y eventos.

Tabla 2.3. *Indicadores y ejemplos de diseño y organización*

<i>Indicadores</i>	<i>Ejemplos</i>
Fijar el programa de estudios	«Esta semana hablaremos de...»
Diseñar métodos	«Voy a repartiros en diferentes grupos y comentaréis... »
Establecer un calendario	«Envía un mensaje antes del viernes, por favor»
Emplear el medio de forma efectiva	«Cuando contestes (por e-mail), intenta referirte a temas que otros hayan planteado»
Establecer pautas de conducta y cortesía en la comunicación electrónica (<i>netiquette</i>)	«Intenta escribir mensajes breves»
Plantear observaciones en el nivel macro del contenido de los cursos	«Esta presentación pretende ofreceros una gama amplia de herramientas y competencias que podréis utilizar para decidir cuándo y cómo emplear las f diferentes técnicas de

	investigación»
--	----------------

Fuente: Garrison y Anderson (2005, p. 100)

2.8.2. FACILITAR EL DISCURSO

El segundo elemento de la presencia docente, facilitar el discurso con el objetivo de construir conocimiento, apunta al centro de la experiencia de e-learning. El hecho de facilitar el discurso implica reconocer el rol de la comunidad de investigación como potenciadora de la construcción de significados así como su papel en la promoción de la comprensión mutua. “Este elemento representa la fusión de objetivos, proceso y resultados. Es el punto en que convergen el interés, el compromiso y el aprendizaje” (Garrison y Anderson, 2005, p. 100).

El profesor desempeña una función clave en el momento de facilitar el discurso en un contexto de e-learning. Gestionar el discurso en ese entorno no es menos importante que en un contexto presencial. La naturaleza reflexiva de la comunicación de base textual exige una actitud seria. Mantener esa actitud y promover las aportaciones de calidad requiere que el discurso esté bien centrado y que sea productivo.

Las responsabilidades docentes requieren una atención continua respecto a un amplio abanico de asuntos. La preocupación central es establecer y mantener una comunidad de estudio. Ello requiere atención hacia los aspectos de presencia cognitiva y social. La comunicación debe ser seguida muy de cerca y prestar mucha atención a la naturaleza y ritmo de las respuestas. La comunidad debe ser, en cierta forma, auto sostenible; por tanto, demasiada presencia docente o demasiado poca puede afectar negativamente al discurso y al proceso de comprensión. Además de mantener ese equilibrio, las comunicaciones procedentes de los profesores deben servir de modelo respecto a la calidad del discurso y configurarlo en vistas a obtener resultados constructivos. También se precisa de mucha orientación para implicar a los estudiantes menos responsables y para evitar que el debate sea dominado siempre por las mismas personas. Estas cuestiones no implican habilidades diferentes a las necesarias para animar un debate presencial (Garrison y Anderson, 2005).

El profesor no debe restringirse a fomentar o valorar las respuestas prolíficas, sino que debe dar ejemplos de respuestas apropiadas y relevantes, llamar la atención sobre las respuestas bien razonadas y establecer asociaciones entre los mensajes. Los participantes deben sentir que el discurso avanza en un sentido concreto y según una determinada pauta temporal. En su

momento, los hilos del discurso deben ser unidos y debe declararse explícitamente el conocimiento compartido.

Facilitar el discurso con el objetivo de construir conocimiento implica aspectos pedagógicos, interpersonales y organizativos. La presencia docente debe relacionarse tanto con el desarrollo cognitivo como con un entorno positivo de aprendizaje, y debe contemplar los contenidos, la cognición y el contexto como partes integrantes del todo. Algunos mensajes son básicamente sociales, tienen lugar en los *chats* y suelen estar fuera del alcance de los profesores. Otros son más complejos e incorporan varios elementos cognitivos y sociales. Aquí es donde entran en juego las capacidades de un profesor y animador.

Tabla 2.4. *Indicadores y ejemplos de facilitar el discurso*

<i>Indicadores</i>	<i>Ejemplos</i>
Identificar áreas de acuerdo/desacuerdo	«Juan, María ha aportado un contra ejemplo fuerte a tu hipótesis. ¿Qué le contestarías?»
Intentar alcanzar un consenso	«Creo que, de hecho, Juan y Maria están diciendo lo mismo»
Animar, reconocer o reforzar las contribuciones de los estudiantes	«Gracias por tus observaciones, son muy acertadas»
Establecer un clima de estudio	«No te preocupes demasiado por evitar «pensar en voz alta» en el forum. Después de todo, este es un lugar para ensayar ideas»
Extraer opiniones de los participantes, el debate	«¿Algún comentario más sobre este asunto?» promover ¿Alguien quiere decir algo más sobre esto? »
Evaluar la eficacia del proceso	«Creo que aquí nos estamos desviando un poco»

Fuente: Garrison y Anderson (2005, p. 102)

2.8.3. ENSEÑANZA DIRECTA

Para Garrison y Anderson (2005), la enseñanza directa va más allá de la función asociada con promover el debate y la participación y suele asociarse más bien con asuntos específicos de contenidos, como el diagnóstico de errores de concepto. El liderazgo del profesor se manifiesta en este tipo de situaciones y suele ser de carácter bastante específico. Aunque se trata de una influencia de enorme importancia, esta responsabilidad docente básica ha sido muchas veces ignorada o dejada de lado. La competencia en asuntos de disciplina y en la configuración eficaz de la experiencia educativa son aspectos esenciales de la misma. El

riesgo inherente al e-learning es que el clima educativo e intelectual se pierda y que los integrantes de la comunidad no respeten las normas.

Los autores destacan que el potencial de la enseñanza directa supone un reto para el concepto de guía paralela. Aunque este enfoque puede ser de cierto valor en algunas ocasiones, supone ciertas limitaciones para el e-learning. Implica una separación artificial entre el facilitador y el experto en contenidos, y puede llegar a distorsionar la experiencia educativa al centrarla exclusivamente en el estudiante dejando al margen la influencia de un experto en pedagogía y contenidos bajo la forma de profesor. Un enfoque demasiado libre lleva a malinterpretar el enfoque colaborativo-constructivista del aprendizaje y la importancia de construir sistemáticamente experiencias educativas para conseguir los resultados esperados.

La presencia docente, tal y como se define aquí, no es posible sin la competencia profesional de un profesor responsable y con experiencia que pueda identificar las ideas y conceptos dignos de estudio, presentar el orden de los conceptos, organizar las actividades educativas, guiar el discurso y ofrecer fuentes adicionales de información, diagnosticar los errores de concepto e intervenir cuando sea necesario. Se trata de intervenciones directas y proactivas que dan soporte a una experiencia educativa eficaz y eficiente (Garrison y Anderson, 2005, p. 104). La tabla 2.5 muestra indicadores y ejemplos de la enseñanza directa.

Tabla 2.5. *Indicadores y ejemplos de enseñanza directa*

<i>Indicadores</i>	<i>Ejemplos</i>
Presentar contenidos/cuestiones	«B. dice... ¿qué pensáis?»
Centrar el debate en temas específicos	«Creo que eso es un callejón sin salida. Te pediría que consideraras...»
Resumir el debate	«La cuestión original era... J. dijo... M. dijo... hemos legado a la conclusión de que... Aún no hemos tocado el tema relativo a...»
Confirmar lo que se ha entendido mediante la evaluación y el feedback explicativo.	«Estás cerca, pero no has dado cuenta de... y eso es importante porque... »
Diagnosticas los errores de concepto	«Recuerda, B. habla desde una perspectiva administrativa, por tanto, ten cuidado cuando dices ...»
Inyectar conocimiento desde diferentes fuentes; por ejemplo libros de texto, artículos, Internet, experiencias personales	«Estuve en una lectura con B. una vez, y dijo ... Puedes encontrar las actas en http://www... »
Responder a las preocupaciones técnicas	«Si quieres incluir un hyperlink en tu mensaje

	tienes que...»
--	----------------

Fuente: Garrison y Anderson (2005, p. 103)

2.9. VISIÓN HOLÍSTICA

Aunque una presencia social fuerte aporta la base necesaria para la crítica respetuosa y constructiva, no garantiza, por sí sola, el funcionamiento óptimo de la comunidad de investigación. Deben darse, al mismo tiempo, los demás elementos de esa comunidad las presencias cognitiva y docente para conseguir el nivel óptimo de presencia social que se necesita para conseguir los objetivos educativos. Las presencias cognitiva y social son los elementos que convierten una comunidad con función social en sentido amplio en una comunidad de investigación

El marco de la presencia cognitiva permite apreciar a fondo los aspectos cognitivos de la experiencia del e-learning y se convierte en un medio para evaluar la naturaleza cualitativa de ese discurso. A la vez, evaluar la naturaleza del discurso puede ayudar a profundizar en la transacción enseñanza-aprendizaje y a descubrir qué intervenciones pueden ser más apropiadas.

Las categorías de la presencia docente ofrecen una plantilla que puede ser de bastante valor para estructurar, facilitar y dirigir una experiencia de e-learning. A pesar de la función esencial del profesor, debe subrayarse que en una comunidad de investigación, marco de esta investigación, todos los participantes tienen la oportunidad de contribuir a la presencia docente. De hecho, si el último objetivo es aprender a aprender, debe animarse a los estudiantes a que sean capaces de autoorientarse y de gestionar su propio estudio. El concepto es presencia *docente* no presencia del *docente*. A medida que los participantes se desarrollan desde el punto de vista cognitivo y social, la presencia docente se vuelve más distribuida.

Por lo tanto, el reto consiste en proponer actividades que capitalicen las características y el potencial del medio en uso. Ello significa pensar de modo diferente sobre lo que puede ser una experiencia de e-learning. Implica reconocer la posibilidad de crear una comunidad de investigación cooperativa y crítica y que sea rica y accesible. Todo ello requiere concebir el rol del profesor no como una presencia reducida sino diferente. Aunque se mantienen muchas

de las responsabilidades docentes, la mayoría de ellas se manifestarán de forma distinta.

2.10. CALIDAD EN ENTORNOS EDUCATIVOS VIRTUALES

2.10.1. Calidad y e-learning

Las plataformas educativas exigen la introducción de cambios sustanciales que fundamenten las teorías y prácticas pedagógicas. En un entorno de e-learning se modifican los modos de interacción docente-alumno mediatizados por el conocimiento y por la finalidad de la enseñanza. Los estudiantes adquieren un protagonismo que les exige reflexionar, producir nuevos conocimientos y utilizarlos (Aballay, Carestia y Martín, 2007).

El objetivo de una educación de calidad está presente en la misión de las instituciones educativas de nivel superior y se instrumenta en sus normas, tal como con la Ley N° 24521 de Educación Superior. Garantizar la calidad en educación es una tarea compleja que exige cambios profundos de orden cultural e institucional.

“La calidad puede ser entendida como el conjunto de propiedades y características de un bien o servicio, que le confiere la aptitud para satisfacer las necesidades y expectativas explícitas o implícitas preestablecidas” (Aballay et al., 2007, p. 2).

Los modelos de calidad siguen lineamientos aceptados a nivel internacional plasmados en estándares y reunidos en normas de calidad, guías reconocidas para la aceptación de bienes y servicios, en los diferentes países del mundo.

La norma internacional ISO 9001:2000 promueve la adopción de un enfoque a procesos dentro de una organización, lo que implica la identificación y gestión de procesos y sus interacciones. Estos estándares permiten que la gestión de la calidad sea aplicable a organizaciones dedicadas a la producción de servicios educativos en sus diferentes modalidades.

De una manera simplificada, en un sistema de e-learning se podrían distinguir dos tipos de procesos. Los *procesos principales* que son el aprendizaje y la investigación para la

producción de nuevos conocimientos y los *procesos de soporte* como la producción de materiales, la gestión de tutoría, la administración y mantenimiento de las plataformas educativas, entre otros.

De cada proceso se obtienen productos. Los *productos obtenidos del proceso principal* no son bienes tangibles, los indicadores de presencia cognitiva son un medio que ayuda a evaluar estos productos. Los productos obtenidos de los procesos de soporte se pueden evaluar con los indicadores de la presencia docente o social. Por ejemplo los materiales con la mediación pedagógica y comunicacional necesarias para la formación no presencial, docentes-tutores con las competencias adecuadas a la modalidad, plataformas que den soporte a la propuesta que se pretende implementar.

Si bien los productos utilizados para e-learning son *productos software*, para los cuales es posible aplicar los estándares de calidad de la Ingeniería de Software, es necesario priorizar la obtención de calidad en cuanto a la función educativa de los mismos. Por esta razón este trabajo se enfoca en la gestión de la calidad relacionada con el proceso de enseñanza-aprendizaje.

2.10.2. El ciclo de Calidad

En Aballay et al. (2007) se expresa:

El ciclo de Deming o ciclo de calidad propone una forma de acción constante en pos del mejoramiento continuo. Plantea cuatro etapas: planear, hacer, verificar y actuar (PHVA). *Planear* incluye determinar metas, objetivos y métodos. *Hacer* implica dar educación y capacitación a los involucrados y realizar el trabajo en sí. *Verificar* consiste en evaluar los resultados de lo hecho. *Actuar* implica tomar medidas correctivas. (p. 3)

El e-learning está convirtiéndose en una de las tendencias de más rápido crecimiento en la Educación Superior. Comparado con la educación tradicional este tipo de aprendizaje requiere mayor disciplina y madurez por parte de los estudiantes, mayor esfuerzo en el planeamiento, diseño, implementación, evaluación y reajustes por parte de los docentes, si se pretenden implementar propuestas pedagógicas de calidad en los ambientes de aprendizaje basados en la red.

Planear en e-learning consiste en establecer las *metas y objetivos* que expresan las necesidades que se han de satisfacer como resultado del esfuerzo educativo y deben

formularse en relación con los indicadores de las distintas presencias, cognitiva, docente y social que se quieren lograr. Estas *metas y objetivos* deben expresarse por escrito, ser concretos, explícitos, suficientemente informativos y distribuirse ampliamente a quienes van dirigidos, sean éstos expertos en contenidos de la disciplina, especialistas y técnicos en la producción de materiales didácticos, tutores/consultores, profesionales informáticos, administrativos o directivos (Aballay et al., 2007).

“Las *metas y objetivos* deben acompañarse con *métodos* que establezcan un modo operativo de alcanzarlos” (Aballay et al., 2007, p. 3). Las categorías de la presencia docente ofrecen una guía que puede utilizarse para estructurar, facilitar y dirigir una experiencia de e-learning. Las categorías de la presencia cognitiva permiten el estudio de los logros a alcanzar, los cambios que deben producirse y los procedimientos de evaluación. Las categorías de la presencia social aportan la base necesaria para el funcionamiento óptimo de la comunidad de investigación. En e-learning el proyecto educativo deberá planificar la adecuada integración de todas las presencias para que se capitalicen las características y el potencial del medio en uso.

Hacer, en e-learning consiste en *capacitar* a los docentes y realizar el proceso de enseñanza-aprendizaje.

Todos los docentes involucrados en esta modalidad deben recibir capacitación actualizada de manera constante y preferentemente a distancia. De esta forma, *poniéndose en el lugar del otro* los docentes pueden vivenciar y sentir como suyas las dificultades, percepciones, frustraciones, avances y logros que experimentarán cognitivamente los estudiantes, destinatarios reales del proceso de enseñanza-aprendizaje. (Aballay et al., 2007, p. 3)

La tercera etapa es *verificar*, en el área educativa se utiliza la evaluación en las diferentes instancias y niveles institucionales como forma de verificar los resultados alcanzados con relación a los objetivos planteados. La evaluación de calidad en un sistema de e-learning se distingue por su enfoque holístico y tiene presente el conjunto total de los elementos que participan en él (Aballay et al., 2007).

En este punto, se considera importante hacer la siguiente aclaración: si bien la evaluación arroja información de suma importancia sobre todos los recursos intervinientes en un modelo psicopedagógico de e-learning -sean éstos recursos humanos, materiales, tecnológicos,

comunicacionales, etc.- y sirve de fundamento a los fines de tomar acciones correctivas, no menos importante resulta *anticiparse y tomar acciones de carácter preventivo* (Aballay et al., 2007).

En la cuarta fase del ciclo, *actuar*, se plantea que a partir del diagnóstico de la situación detectada en la verificación se deben investigar las causas, establecer medidas correctivas y ponerlas en acción. En e-learning, resulta prioritario emprender acciones apropiadas que apunten a corregir, ajustar, reformar, desarrollar, o fortalecer aquellas áreas que inciden en la calidad educativa (Aballay et al., 2007).

En el marco de un sistema de e-learning, la retroalimentación que se logra a partir del seguimiento, el análisis y evaluación de debilidades y fortalezas debe ser utilizada para formular planes de mejoras y convertir la gestión de la calidad en un objetivo decisivo del sistema (Aballay et al., 2007).

Las técnicas de minería de datos pueden ayudar a la comprensión del e-learning y permitir a los docentes descubrir conocimiento para implementar estrategias de enseñanza-aprendizaje que permitan la mejora continua del proceso.

CAPÍTULO 3

Minería de Datos

CAPÍTULO 3: MINERÍA DE DATOS

3.1. INTRODUCCIÓN

La cantidad y variedad de datos que se encuentran almacenados en forma digital crece vertiginosamente desde hace varias décadas. Gran parte de estos datos se generan en los sistemas que registran las operaciones o transacciones más importantes de las organizaciones como son ventas, compras, pagos, cobros, extracciones, depósitos, etc. Su función de memoria de la organización es útil para explicar el pasado, entender el presente y predecir la información futura. La mayoría de las decisiones en las organizaciones se basan en esa información.

El análisis e interpretación de los datos realizados en forma manual es la forma tradicional en se puede obtener información y transformarla en conocimiento para contribuir al proceso de toma de decisiones. Esta forma de actuar es lenta, cara y altamente subjetiva. Además, el análisis manual es impracticable en dominios donde el volumen de los datos desborda la capacidad humana de comprenderlos. Por este motivo muchas decisiones importantes se realizan siguiendo la intuición del usuario y no se utiliza el conocimiento contenido en la gran cantidad de datos disponibles. Este es el principal cometido de la minería de datos: resolver problemas analizando los datos presentes en las bases de datos (Hernández et al., 2004).

Ha surgido una nueva generación de herramientas y técnicas para soportar la extracción de conocimiento útil desde la información disponible, y que se engloban bajo la denominación de *minería de datos* o *data mining*. El resultado de la minería de datos son conjuntos de reglas, ecuaciones, árboles de decisión, redes neuronales, grafos probabilísticas, los cuales pueden usarse para el apoyo a la toma de decisiones en las organizaciones.

3.2. DEFINICIÓN DE MINERÍA DE DATOS

Witten y Frank (2000) definen la minería de datos como “el proceso de extraer conocimiento útil y comprensible, previamente desconocido, desde grandes cantidades de datos almacenados en distintos formatos” (p. 5). Es decir, la tarea fundamental de la minería de

datos es encontrar modelos inteligibles a partir de los datos. Para que este proceso sea efectivo debería ser automático o semiautomático y el uso de los patrones descubiertos deberá ayudar a tomar decisiones más seguras que reporten, por lo tanto, algún beneficio a la organización (Hernández et al., 2004).

La utilidad del conocimiento extraído está íntimamente relacionada con la comprensibilidad del modelo inferido. Generalmente el usuario final no es un experto en las técnicas de minería de datos, ni tampoco puede perder mucho tiempo en interpretar los resultados. Por esto, en muchas aplicaciones es importante hacer que la información descubierta sea más comprensible utilizando representaciones gráficas, convirtiendo los patrones a lenguaje natural o utilizando técnicas de visualización de los datos.

De una manera simple pero ambiciosa, se puede decir que el objetivo de la minería de datos es convertir datos en conocimiento. Se enumeran algunas de las aplicaciones más conocidas de la minería de datos:

Aplicaciones financieras y bancarias

- Obtención de patrones de uso fraudulento de tarjetas de crédito.
- Determinación del gasto en tarjeta de crédito por grupos.
- Análisis de riesgo de créditos.

Análisis de mercado, distribución y comercio en general

- Análisis de la cesta de compras. Compras conjuntas, secuenciales, ventas cruzadas, señuelos, etc.
- Evaluación de campañas publicitarias.
- Análisis de fidelidad de clientes. Reducción de fuga.
- Segmentación de clientes.

Seguros y salud privada

- Determinación de los clientes que podrían ser potencialmente caros.
- Análisis de procedimientos médicos solicitados conjuntamente.
- Identificación de patrones de comportamiento de clientes con riesgo.
- Educación
- Selección o captación de estudiantes.
- Detección de abandono y de fracaso.

- Estimación del tiempo de estancia en la institución.

3.3. TIPOS DE MODELOS

El conocimiento generado por la minería de datos puede expresarse en forma de relaciones, patrones o reglas, o bien en forma de un resumen de los datos. Estas relaciones o resúmenes constituyen el modelo de los datos analizados. Existen muchas formas diferentes de representar los modelos y cada una de ellas determina el tipo de técnica que puede usarse para inferirlos (Hernández et al., 2004).

Los modelos pueden ser de dos tipos: *predictivos* o *descriptivos*. “Los modelos predictivos pretenden estimar valores futuros o desconocidos de variables de interés, que se denominan variables objetivo o dependientes, usando otras variables o campos de la base de datos, denominadas variables independientes o predictivas” (Hernández et al., 2004, p. 12). Por ejemplo, un modelo predictivo sería aquel que permite identificar que clientes podrían no devolver un préstamo solicitado, teniendo en cuenta la información sobre préstamos anteriores y datos personales del mismo cliente.

“Los modelos descriptivos, en cambio, identifican patrones que explican o resumen los datos, es decir, sirven para explorar las propiedades de los datos examinados, no para predecir nuevos datos” (Hernández et al., 2004, p. 12). Por ejemplo un supermercado desea obtener información sobre los hábitos de compra de sus clientes, con el objetivo de mejorar la ubicación de los productos en el local.

Algunas tareas de minería de datos que producen modelos predictivos son la clasificación y la regresión, y las que dan lugar a modelos descriptivos son el agrupamiento, las reglas de asociación y el análisis correlacional. Cada tarea puede ser realizada usando distintas técnicas. Por ejemplo, los modelos inferidos por los árboles de decisión y las redes neuronales pueden inferir modelos predictivos. Igualmente, para una misma técnica se han desarrollado diferentes algoritmos que difieren en la forma y criterios concretos con los que se construye el modelo.

En este trabajo se describen dos técnicas, una predictiva y otra descriptiva que se utilizan para el descubrimiento de conocimiento en los foros de aprendizaje de una propuesta de e-learning.

3.4. EL PROCESO DE DESCUBRIMIENTO DE CONOCIMIENTO

Existen términos que se utilizan frecuentemente como sinónimos de *minería de datos*. Uno de ellos se conoce como *análisis inteligente de datos*, otro muy utilizado es *descubrimiento de conocimiento en bases de datos* o *Knowledge Discovery in Database* o *KDD* por sus siglas en inglés.

Fayyad and Stolorz (1997, p. 99) definen el KDD como “el proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y, en última instancia, comprensibles a partir de los dato”. En esta definición se resumen cuáles deben ser las propiedades deseables del conocimiento extraído:

- válido: hace referencia a que los patrones deben seguir siendo precisos para datos nuevos, con un cierto grado de certidumbre, y no sólo aquellos que han sido usados en su obtención.
- novedoso: que aporte algo novedoso tanto para el sistema y como para el usuario.
- potencialmente útil: la información debe conducir a acciones que reporten algún tipo de beneficio para el usuario.
- comprensible: la extracción de patrones no comprensibles dificulta o imposibilita su interpretación, revisión, validación y uso en la toma de decisiones. De hecho, una información incomprensible no proporciona conocimiento, al menos desde el punto de vista de su utilidad.

Cómo se deduce de la definición anterior, KDD es un proceso complejo que incluye varias fases una de ellas es la minería de datos. Los sistemas KDD permiten la selección, limpieza, transformación y proyección de los datos; analizar los datos para extraer patrones y modelos adecuados; evaluar e interpretar los patrones para convertirlos en conocimiento; consolidar el conocimiento resolviendo posibles conocimientos previamente extraído; y hacer el conocimiento disponible para su uso. Esta definición del proceso clarifica la relación entre el KDD y la minería de datos: el KDD es el proceso global de descubrir conocimiento útil desde las bases de datos mientras que la minería de datos se refiere a la aplicación de los métodos de

aprendizaje y estadísticos para la obtención de patrones y modelos. Al ser la fase de generación de modelos, comúnmente se asimila KDD con minería de datos.

3.5. REGLAS DE ASOCIACIÓN

Hernández et al. (2004) expresan que:

Las reglas de asociación son una manera muy popular de expresar patrones de datos de una base de datos. Estos patrones pueden servir para conocer el comportamiento general del problema que genera la base de datos, y de esta manera tener más información que pueda asistir en la toma de decisiones. (p. 237)

Las reglas de asociación surgieron inicialmente para afrontar el análisis de las cestas de la compra de los comercios. Las diferentes cestas de compras se pueden expresar en una tabla. Las filas de la tabla se refieren a una cesta de un supermercado y las columnas son cada uno de los productos en venta en el supermercado. La tabla solo contiene valores binarios. Un 1 en la posición (i,j) indica que la cesta i incorpora el producto j , mientras que un 0 indica que el cliente no ha adquirido el producto. La tabla 3.1 se da como ejemplo de una base de datos de este tipo.

Formalmente, sea $I = \{i_1, i_2, i_3, \dots, i_m\}$ un conjunto de ítems,

sea $T = \{t_1, t_2, t_3, \dots, t_n\}$ un conjunto de transacciones,

donde cada transacción t_i es un conjunto de ítems tal que $t_i \subseteq I; 1 \leq i \leq n$

La implicación $X \Rightarrow Y$ es una regla de asociación donde $X \subset I, Y \subset I$,

$X \cap Y = \emptyset$, y $X \cup Y \subseteq t_i$ esto es, el conjunto de ítems formado por aquellos que corresponden al antecedente o al consecuente de la regla de asociación, debe estar contenido o ser igual a alguna de las transacciones pertenecientes a T .

Tabla 3.1 Base de datos con 4 productos y 5 transacciones

ID	Leche La vaca	GalletasTriguillas	Manteca La Blanca	Gaseosa Dulcinea
t1	1	1	0	0
t2	0	1	1	0
t3	0	0	0	1
t4	1	1	1	0
t5	0	1	0	0

Para ilustrar estos conceptos véase el siguiente ejemplo sobre ventas en un supermercado. El conjunto I de ítems o productos es:

$I = \{\text{Leche La Vaca, Galletas Triguillas, Manteca La Blanca, Gaseosa Dulcinea}\}$

La tabla 3.1 se muestra una pequeña base de datos que contiene los productos, donde el código '1' se interpreta como que el producto correspondiente está presente en la transacción y el código '0' significa que dicho producto no está presente. Un ejemplo de regla para el supermercado podría ser:

$$\{\text{Leche La Vaca, Galletas Triguillas}\} \Rightarrow \{\text{Manteca La Blanca}\}$$

Significaría que si el cliente compró Leche La Vaca y Galletas Triguillas también compró Manteca La Blanca, es decir, según la especificación formal anterior se tendría que:

$$X = \{\text{Leche La Vaca, Galletas Triguillas}\}$$

$$Y = \{\text{Manteca La Blanca}\}$$

$$X \Rightarrow Y$$

Dada una regla de asociación, se acostumbra a trabajar con dos medidas para conocer la calidad de la regla: *soporte* y *confianza*. En inglés *support* y *confidence*. El soporte, también denominado cobertura de una regla se define como el número de instancias que la regla predice correctamente. Es decir, el soporte puede ser considerado como la probabilidad de que las transacciones contengan un conjunto de ítems. Por otra parte la confianza, también llamada precisión, mide el porcentaje de veces que la regla se cumple cuando se puede aplicar. Es decir, es la probabilidad de que las transacciones que contienen el antecedente de la regla, también contengan al consecuente. La confianza puede ser considerada, entonces, como un estimador de la probabilidad condicional $p(x/y)$ (Hernández et al., 2004).

En el ejemplo anterior, el conjunto $\{\text{Leche La Vaca, Galletas Triguillas}\}$ tiene un soporte:

$$\text{sop}(X) = \frac{2}{5} = 0.4$$

Es decir, hay dos transacciones o instancias en que el conjunto X está presente, de un total de cinco transacciones, por lo tanto el soporte es del 40%.

La confianza sería.

$$X \cup Y = \{\text{Leche La Vaca, Galletas Triguillas, Manteca La Blanca}\}$$

$$\text{sop}(X \cup Y) = 0,2$$

$$\text{conf}(X \Rightarrow Y) = \frac{\text{sop}(X \cup Y)}{\text{sop}(X)} = \frac{0.2}{0.4} = 0.5$$

Este cálculo significa que el 50% de las reglas de la base de datos que contienen en el antecedente *Leche La Vaca* y *Galletas Triguillas* tienen *Manteca La Blanca* en el consecuente; en otras palabras, que la regla:

{Leche La Vaca, Galletas Triguillas} => {Manteca La Blanca} es cierta en el 50% de los casos.

Normalmente, el aprendizaje de reglas de asociación se basa en su soporte y su confianza. Los algoritmos de aprendizaje trabajan en la búsqueda de reglas que cumplan requisitos mínimos de estas medidas. Dado el gran volumen de datos de los problemas con el que los algoritmos de aprendizaje de reglas de asociación trabajan, la tarea de buscar patrones que cumplan estos requisitos puede ser costosa, ya que los conjuntos de ítems a ser analizados crece exponencialmente con respecto al número de variables de los datos. Sin embargo, en los casos reales existen realmente pocos conjuntos frecuentes y los métodos que exigen una confianza y una cobertura mínima se benefician de este hecho. Por ejemplo, la mayoría de los clientes de un supermercado suele comprar un número muy limitado de productos.

Las restricciones que pueden presentar los factores de soporte y confianza, unido a la necesidad de rescatar reglas interesantes de un conjunto posiblemente numeroso de reglas generadas, dificulta la actividad de los expertos en el campo de la minería. A raíz de ello, distintos autores han desarrollado diferentes medidas para evaluar la importancia de las reglas generadas. Entre esas medidas se rescata una muy popular, denominada *lift*, interés o medida de independencia. El factor lift representa un test para medir la independencia estadística, y se define como:

$$\text{lift}(X \Rightarrow Y) = \frac{\text{conf}(X \Rightarrow Y)}{\text{sop}(Y)}$$

Lift establece una relación entre la ocurrencia simultánea de X e Y en las transacciones, respecto a que los conjuntos de ítems que conforman el antecedente y el consecuente de la regla sean estadísticamente independientes. Un valor de lift mayor a 1 indica que la regla es

buena, que los ítems están asociados. Valores de lift igual o menor a 1 pueden indicar la independencia de los ítems. Se aconseja que las reglas recuperadas por medio de soporte y confianza sean filtradas usando su valor lift (Hernández et al., 2004).

3.6. ÁRBOLES DE DECISIÓN

Según los autores Hernández et al. (2004):

Un árbol de decisión es un conjunto de condiciones organizadas en una estructura jerárquica, de tal manera que la decisión final a tomar se puede determinar siguiendo las condiciones que se cumplen desde la raíz del árbol hasta alguna de sus hojas. (p. 282)

Los árboles de decisión se utilizan desde hace siglos, y son especialmente apropiados para expresar procedimientos médicos, legales, comerciales, estratégicos, matemáticos, lógicos, etc.

Una de las grandes ventajas de los árboles de decisión es que, en su forma más general, las opciones posibles a partir de una determinada condición son excluyentes. Esto permite analizar una situación y, siguiendo el árbol de decisión apropiadamente, llegar a una sola acción o decisión a tomar.

La tarea de aprendizaje para la cual los árboles de decisión se adecuan mejor es la clasificación. Clasificar es determinar entre varias clases a que clase pertenece un objeto, la estructura de condición y ramificación de un árbol de decisión es idónea para este problema (Hernández et al., 2004).

Debido al hecho de que la clasificación trata con clases o etiquetas disjuntas, un árbol de decisión conducirá un ejemplo hasta una y sólo una hoja, asignando, por tanto, una única clase al ejemplo. Para ello, las particiones existentes en el árbol deben ser también disjuntas. Es decir cada instancia cumple o no cumple una condición. Además, dicha propiedad es exhaustiva, es decir, una de las condiciones se debe cumplir.

Las particiones son un conjunto de condiciones exhaustivas y excluyentes. Cuantos más tipos de condiciones se permiten, más posibilidades se tendrán de encontrar los patrones que hay

detrás de los datos. Es decir, mientras más particiones se eligen más expresivos y más precisos podrán ser los árboles, pero la complejidad será mayor. Por lo tanto un buen algoritmo de aprendizaje de árboles de decisión es aquel que permite un buen compromiso entre expresividad y eficiencia. Existen numerosos algoritmos o sistemas de aprendizaje de árboles de decisión se describen algunos de los más importantes.

3.6.1. Algoritmo ID3

En 1986 R. J. Quinlan describe un algoritmo de inducción de árboles de decisión (Induction of Decision Trees) identificado por la sigla *ID3*. El sistema ID3 es un algoritmo simple y, sin embargo, potente, cuya misión es la elaboración de un árbol de decisión y pertenece a la familia Top-Down Induction of Decision Trees (TDIDT) que aplican la inducción de arriba hacia abajo en los árboles de decisión. El procedimiento para generar un árbol de decisión consiste, en seleccionar un atributo como raíz del árbol y crear una rama con cada uno de los posibles valores de dicho atributo. Con cada rama resultante (nuevo nodo del árbol), se realiza el mismo proceso, esto es, se selecciona otro atributo y se genera una nueva rama para cada posible valor del atributo. Este procedimiento continúa hasta que los ejemplos se clasifiquen a través de uno de los caminos del árbol. El nodo final de cada camino será un nodo hoja, al que se le asignará la clase correspondiente. Así, el objetivo de los árboles de decisión es obtener reglas o relaciones que permitan clasificar a partir de los atributos.

En cada nodo del árbol de decisión se debe seleccionar un atributo para seguir dividiendo, y el criterio que se toma para elegirlo es: se selecciona el atributo que mejor separe (ordene) los ejemplos de acuerdo a las clases. Para ello se emplea la entropía, que es una medida de cómo está ordenado el universo. La teoría de la información (basada en la entropía) calcula el número de bits (información, preguntas sobre atributos) que hace falta suministrar para conocer la clase a la que pertenece un ejemplo. Cuanto menor sea el valor de la entropía, menor será la incertidumbre y más útil será el atributo para la clasificación.

La formulación matemática que la entropía (H) tiene es la siguiente:

$$H = \sum_{i=1}^n p_i * \log(1 / p_i)$$

Donde: Las p_i son las probabilidades de ocurrencia de los distintos elementos, \log es el logaritmo, en la base que se pacte. Si se usa *log en base 2*, la unidad en que se mida esa entropía será bits de información y la sumatoria es sobre todos los elementos de la fuente, que se asume como n .

En la tabla 3.2 se resume el conocimiento existente de la observación por varios años de las condiciones meteorológicas para las mañanas de un día domingo, durante un largo período de tiempo, así por ejemplo un día domingo se puede presentar con las siguientes características:

tiempo: claro

temperatura: agradable

humedad: normal

viento: suave

Para formar un objeto resta definir un resultado de este objeto que posee un conjunto de valores observados de acuerdo con un dominio de conocimiento que se quiere adquirir. Esta observación podría ser, en este caso, dos atributos ya definidos, como ser quedarse en casa o practicar un deporte. De esta manera el resultado tendrá los valores casa, básquet y fútbol.

Tabla 3.2 Resumen de condiciones meteorológicas para las mañanas del día domingo

Nº	TIEMPO	TEMPERATURA	HUMEDAD	VIENTO	RESULTADO
01	CLARO	AGRADABLE	NORMAL	FUERTE	BASQUET
02	NUBLADO	FRIO	ALTA	SUAVE	CASA
03	NUBLADO	AGRADABLE	ALTA	SUAVE	BASQUET
04	LLUVIOSO	FRIO	ALTA	FUERTE	CASA
05	NUBLADO	CALUROSO	NORMAL	SUAVE	FÚTBOL
06	CLARO	CALUROSO	NORMAL	SUAVE	FÚTBOL
07	NUBLADO	AGRADABLE	ALTA	FUERTE	BASQUET
08	LLUVIOSO	FRIO	ALTA	SUAVE	CASA
09	NUBLADO	AGRADABLE	NORMAL	FUERTE	FÚTBOL
10	CLARO	CALUROSO	ALTA	SUAVE	CASA
11	LLUVIOSO	CALUROSO	ALTA	FUERTE	BASQUET

La raíz del árbol de decisión corresponde al atributo de menor entropía (se le simboliza con la letra H), esto es aquel que presenta menor incertidumbre con respecto a la clase o resultado.

En este ejemplo la entropía de un atributo "x" (Tiempo, Temperatura, Humedad, Viento) es una media ponderada de las entropías de los valores de

$$H(\text{resultado} / \text{atributo "x"}) = \sum_{i=1}^n p_{(vi)} * H(\text{resultado} / \text{valor "vi"})$$

Donde: **n** es el total de valores del atributo "x"; $p(vi)$ es la probabilidad de aparecer ,en el conjunto, de un valor vi . $H(\text{resultado}/vi)$ es la entropía del valor vi (ej. FUERTE, SUAVE) correspondiente al atributo "x" (ej. Viento), y responde a la siguiente expresión:

$$H(\text{resultado} / vi) = \sum_{j=1}^m p(cj / vi) * \log[1 / p(cj / vi)]$$

Donde: **m** es número total de clases o resultados; $p(cj/vi)$ es la probabilidad de tener un ejemplo con valor vi dado que se ha concluido cj . Se calcularán los valores de acuerdo a la Tabla 3.2, atendiendo a la forma de cálculo implementada en ID3. De ahora en adelante se trabaja el logaritmo en base 2.

$$H(\text{resultado=casa/tiempo=nublado})=p(\text{casa/nublado})*\log_2[1/p(\text{casa/nublado})]=1/5\log_2[1/(1/5)]=0,46 \text{ bit de información}$$

El valor de $p(\text{casa/nublado})=1/5$ pues solo en 5 ejemplos el atributo tiempo toma el valor nublado y en solo uno de ellos el resultado es casa.

$$H(\text{resultado=básquet/tiempo=nublado})=p(\text{básquet/nublado})*\log_2[1/p(\text{básquet/nublado})]=2/5\log_2 5/2=0,53 \text{ bit de información}$$

$$H(\text{resultado=fútbol/tiempo=nublado})=p(\text{fútbol/nublado})*\log_2[1/p(\text{fútbol/nublado})]=2/5\log_2 5/2=0,53 \text{ bit de información}$$

Realizando la sumatoria de los valores anteriores se obtiene la entropía para un resultado dado que el tiempo está nublado.

$$H(\text{resultado/tiempo=nublado})=H(\text{resultado=casa/tiempo=nublado})+H(\text{resultado=básquet/tiempo=nublado})+ H(\text{resultado=fútbol/tiempo=nublado}) = 0,46+0,53+0,53=1,52 \text{ bit de información.}$$

Para el resultado dado que el tiempo está lluvioso, quedaría:

$$H(\text{resultado=casa/tiempo=lluvioso})=p(\text{casa/lluvioso})*\log_2[1/p(\text{casa/lluvioso})]=2/3\log_2 3/2=0,389 \text{ bit de información}$$

$$H(\text{resultado}=\text{básquet}/\text{tiempo}=\text{lluvioso})=p(\text{básquet}/\text{lluvioso})\cdot\log_2[1/p(\text{básquet}/\text{lluvioso})]=$$

$$1/3\log_23=0,528 \text{ bit de información}$$

$H(\text{resultado}=\text{fútbol}/\text{tiempo}=\text{lluvioso})=0$ pues $p(\text{fútbol}/\text{lluvioso})=0$ y la función para ese valor de probabilidad tiende a cero

$$H(\text{resultado}/\text{tiempo}=\text{lluvioso})=H(\text{resultado}=\text{casa}/\text{tiempo}=\text{lluvioso})+$$

$$H(\text{resultado}=\text{básquet}/\text{tiempo}=\text{lluvioso})+ H(\text{resultado}=\text{fútbol}/\text{tiempo}=\text{lluvioso}) =$$

$$0,389+0,528+0=0,917 \text{ bit de información.}$$

Para el resultado dado que el tiempo es claro, quedaría:

$$H(\text{resultado}=\text{casa}/\text{tiempo}=\text{claro})=p(\text{casa}/\text{claro})\cdot\log_2[1/p(\text{casa}/\text{claro})]=1/3\log_23= 0,528 \text{ bit de información}$$

$$H(\text{resultado}=\text{básquet}/\text{tiempo}=\text{claro})=p(\text{básquet}/\text{claro})\cdot\log_2[1/p(\text{básquet}/\text{claro})]=$$

$$1/3\log_23=0,528 \text{ bit de información}$$

$$H(\text{resultado}=\text{fútbol}/\text{tiempo}=\text{claro})= p(\text{fútbol}/\text{claro})\cdot\log_2[1/p(\text{fútbol}/\text{claro})]= 1/3\log_23=0,528$$

$$\text{bit de información}$$

$$H(\text{resultado}/\text{tiempo}=\text{claro})=H(\text{resultado}=\text{casa}/\text{tiempo}=\text{claro})+ H(\text{resultado}=\text{básquet}/\text{tiempo}=\text{claro})+$$

$$H(\text{resultado}=\text{fútbol}/\text{tiempo}=\text{claro}) = 0,528+0,528+0,528=1,584 \text{ bit de información}$$

Finalmente el valor de la entropía para el resultado según el atributo tiempo es la suma pesada de las entropías antes calculadas, el peso por el cual se afecta cada entropía no es más que el valor de la probabilidad de ocurrencia de cada valor dentro de ese atributo

$$p(\text{resultado}/\text{tiempo}=\text{casa})=5/11; p(\text{resultado}/\text{tiempo}=\text{lluvioso})=3/11;$$

$$p(\text{resultado}/\text{tiempo}=\text{claro})=3/11;$$

$$H(\text{resultado}/\text{tiempo})=p(\text{resultado}/\text{casa})H(\text{resultado}/\text{casa})+$$

$$p(\text{resultado}/\text{lluvioso})H(\text{resultado}/\text{lluvioso})+ p(\text{resultado}/\text{claro})H(\text{resultado}/\text{claro})=$$

$$5/11*1,52+3/11*0,917+3/11*1,584=1,373 \text{ bit de información}$$

$$H(\text{resultado}/\text{tiempo})=1,373 \text{ bit de información}$$

De la misma forma se calculan las entropías de los otros atributos y surgen los siguiente valores:

$$H(\text{resultado}/\text{temperatura})=0,85 \text{ bit de información}$$

$$H(\text{resultado}/\text{humedad})=0,92 \text{ bit de información}$$

$$H(\text{resultado}/\text{viento})=1,41 \text{ bit de información}$$

Se desprende de ellos que es el atributo *temperatura* es el que menos entropía tiene, esto es sobre el que se tiene menos incertidumbre y éste atributo será la raíz del árbol de conocimiento o decisión, como lo expresa la figura 3.1..

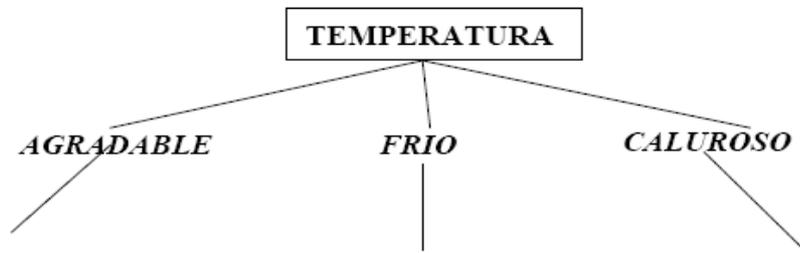


Figura 3.1 Raíz del árbol de decisión

A continuación se debe encontrar la raíz de cada subárbol que comienza en agradable, o en frío o en caluroso; para ello se realizará el mismo procedimiento visto anteriormente. Comenzando por agradable se tendrá la tabla 3.3.

Dado que para tiempo claro

$$p(\text{resultado=fútbol/tiempo=claro}) = p(\text{resultado=casa/tiempo=claro}) = 0$$

su entropía también es nula.

$$H(\text{resultado=básquet/tiempo=claro}) = p(\text{resultado/claro}) * \log[1/p(\text{resultado/claro})] = 0$$

pues queda como resultado del $\log 1 = 0$.

Tabla 3.3. Resumen para temperatura agradable

Nº	TIEMPO	TEMPERATURA	HUMEDAD	VIENTO	RESULTADO
01	CLARO	AGRADABLE	NORMAL	FUERTE	BASQUET
03	NUBLADO	AGRADABLE	ALTA	SUAVE	BASQUET
07	NUBLADO	AGRADABLE	ALTA	FUERTE	BASQUET
09	NUBLADO	AGRADABLE	NORMAL	FUERTE	FÚTBOL

De modo que carece de sentido preguntar por que deporte practicar cuando la temperatura es agradable y el tiempo es claro.

$H(\text{resultado}=\text{básquet}/\text{tiempo}=\text{nublado})=p(\text{básquet}/\text{nublado})\cdot\log[1/p(\text{básquet}/\text{tiempo}=\text{nublado})]=2/3 \log 3/2 = 0,389$ bit de inf.

$H(\text{resultado}=\text{fútbol}/\text{tiempo}=\text{nublado})=p(\text{fútbol}/\text{nublado})\cdot\log[1/p(\text{fútbol}/\text{nublado})]=1/3 \log 3 = 0,528$ bit de información.

Como $p(\text{resultado}=\text{casa}/\text{tiempo}=\text{nublado})=0$ su entropía también lo será.
 $H(\text{resultado}/\text{tiempo}=\text{nublado}) = 0,389 + 0,528 = 0,917$ bit de información

Dado que no existe valor lluvioso para el atributo tiempo no se calcula su entropía

$H(\text{resultado}/\text{tiempo})=p(\text{resultado}/\text{nublado})\cdot H(\text{resultado}/\text{nublado}) = 3/4 * 0.917 = 0,687$ bit de información.

Se calcula ahora la entropía para el atributo humedad; dentro de humedad se observa que para su valor *alto* siempre practico básquet, de modo que no brinda información y su entropía valdrá cero.

$H(\text{resultado}=\text{básquet}/\text{humedad}=\text{normal})=p(\text{básquet}/\text{normal})\cdot\log[1/p(\text{básquet}/\text{normal})]=1/2 \log 2 = 0,5$ bit de información.

$H(\text{resultado}=\text{fútbol}/\text{humedad}=\text{normal})=p(\text{fútbol}/\text{normal})\cdot\log[1/p(\text{fútbol}/\text{normal})]=1/2 \log 2 = 0,5$ bit de información.

$H(\text{resultado}/\text{humedad}=\text{normal}) = H(\text{básquet}/\text{normal}) + H(\text{fútbol}/\text{normal}) = 0,5 + 0,5 = 1$ bit de información.

$H(\text{resultado}/\text{humedad})=p(\text{resultado}/\text{normal})\cdot H(\text{resultado}/\text{normal}) = 1/2 * 1 = 0,5$ bit de información

Se hace el mismo procedimiento para el atributo viento; se observa de la Tabla 3.3. que si el viento es suave y la temperatura agradable, se practica básquet, de modo que no tiene contenido informacional.

$H(\text{resultado}=\text{básquet}/\text{viento}=\text{fuerte})=p(\text{básquet}/\text{fuerte})\cdot\log[1/p(\text{básquet}/\text{fuerte})]=2/3 \log 3/2 = 0,389$ bit de información.

$H(\text{resultado}=\text{fútbol}/\text{viento}=\text{fuerte})=p(\text{fútbol}/\text{fuerte})\cdot\log[1/p(\text{fútbol}/\text{fuerte})]=1/3 \log 3 = 0,528$ bit de información.

$H(\text{resultado}/\text{viento}=\text{fuerte})=H(\text{resultado}=\text{básquet}/\text{viento}=\text{fuerte}) +$

$H(\text{resultado}=\text{fútbol}/\text{viento}=\text{fuerte}) = 0,917$ bit de información.

$H(\text{resultado}/\text{viento}) = p(\text{resultado}/\text{viento}=\text{fuerte})\cdot H(\text{resultado}/\text{viento}=\text{fuerte})=3/4 * 0,917 = 0,687$ bit de información.

Del análisis comparativo de $H(\text{resultado}/\text{viento})=0,687$; $H(\text{resultado}/\text{tiempo})=0,687$; $H(\text{resultado}/\text{humedad})=0,5$ bit de información respectivamente se desprende que de todos los atributos pendientes es el de humedad el que menor incerteza posee, y será por éste atributo que se preguntará a continuación en la rama iniciada con temperatura agradable.

Como conclusión de lo anterior se desprende la tabla 3.4 para temperatura agradable y humedad normal.

Tabla 3.4. Resumen para temperatura agradable y humedad normal

Nº	TIEMPO	TEMPERATURA	HUMEDAD	VIENTO	RESULTADO
01	CLARO	AGRADABLE	NORMAL	FUERTE	BASQUET
09	NUBLADO	AGRADABLE	NORMAL	FUERTE	FÚTBOL

De la Tabla 3.2 surge que es el estado del atributo tiempo el que va indicar que deporte practicar ya que el estado de viento es siempre fuerte; de ésta manera se completa la rama correspondiente a temperatura agradable dando el subárbol de la figura 3.2.

Exactamente el mismo análisis que se hizo para la rama de temperatura agradable debería hacerse para frío y caluroso, éste análisis así indicado da como resultado el árbol de conocimiento *completo* que se observa en la figura 3.3. respecto del universo que originalmente representó la Tabla 3.2.

Del árbol anterior se desprende que el atributo viento no brinda información excluyente como para tenerlo en cuenta. De todos los posibles árboles que se puedan obtener, es éste, el que

menos hojas tiene y mediante el cual se realiza la menor cantidad de preguntas para satisfacer una consulta sobre que actividad desarrollar un día domingo a la mañana.

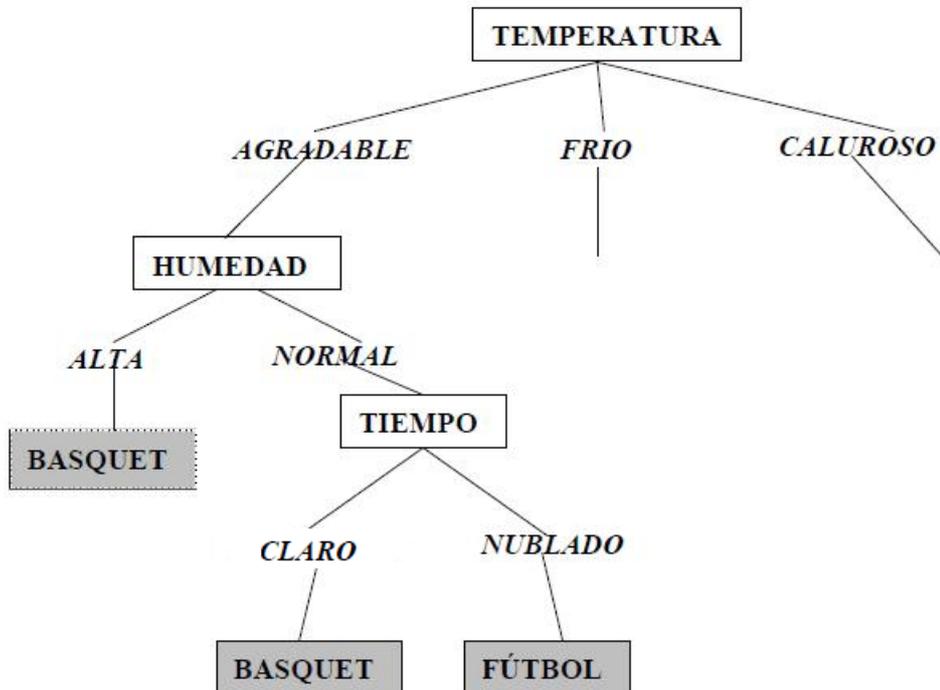


Figura 3.2 Subárbol correspondiente a temperatura agradable

Del árbol anterior se desprende que el atributo viento no brinda información excluyente como para tenerlo en cuenta. De todos los posibles árboles que se puedan obtener, es éste, el que menos hojas tiene y mediante el cual se realiza la menor cantidad de preguntas para satisfacer una consulta sobre que actividad desarrollar un día domingo a la mañana.

Del árbol completo correspondiente a temperatura se obtienen las siguientes reglas, que pueden ser ya almacenadas en una computadora por medio de cualquier lenguaje:

REGLA1

SI (TEMPERATURA=AGRADABLE) AND (HUMEDAD=ALTA) **ENTONCES** (RESULTADO=BASQUET)

REGLA2

SI (TEMPERATURA=AGRADABLE) AND (HUMEDAD=NORMAL) AND (TIEMPO=CLARO) **ENTONCES** (RESULTADO=BASQUET)

REGLA3

SI (TEMPERATURA=*AGRADABLE*) **AND** (HUMEDAD0=*NORMAL*) **AND** (TIEMPO0=*NUBLADO*) **ENTONCES** (RESULTADO=*FÚTBOL*)

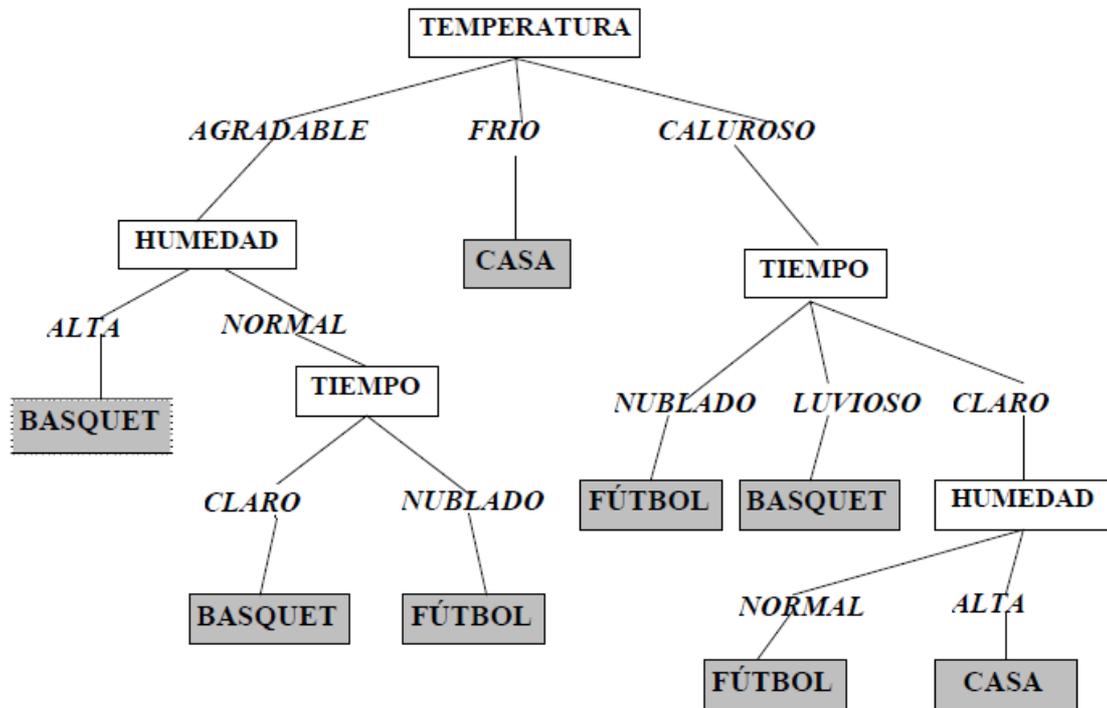


Figura 3.3 Árbol completo correspondiente a temperatura

REGLA4

SI (TEMPERATURA=*FRIO*) **ENTONCES** (RESULTADO=*CASA*)

REGLA5

SI (TEMPERATURA=*CALUROSO*) **AND** (TIEMPO=*NUBLADO*) **ENTONCES** (RESULTADO=*FÚTBOL*)

REGLA6

SI (TEMPERATURA=*CALUROSO*) **AND** (TIEMPO=*LLUVIOSO*) **ENTONCES** (RESULTADO=*BASQUET*)

REGLA7

SI (TEMPERATURA=*CALUROSO*) **AND** (TIEMPO=*CLARO*) **AND** (HUMEDAD=*NORMAL*) **ENTONCES** (RESULTADO=*FÚTBOL*)

REGLA8

SI (TEMPERATURA=AGRADABLE) AND (TIEMPO=CLARO) AND (HUMEDAD=ALTA)
ENTONCES (RESULTADO=CASA)

3.6.2. Weka

En este trabajo se utiliza como herramienta de prueba de los algoritmos de aprendizaje el software *Waikato Environment for Knowledge Analysis* que es un entorno para análisis del conocimiento desarrollado por la Universidad de Waikato, conocido por su sigla en inglés WEKA. La Weka (*Gallirallus australis*) es un ave endémica de Nueva Zelanda. Este ave da nombre a una extensa colección de algoritmos de conocimiento de máquina desarrollados por la Universidad de Waikato e implementados en Java; útiles para ser aplicados sobre datos mediante las interfaces que ofrece o para embeberlos dentro de cualquier aplicación. Además Weka contiene las herramientas necesarias para realizar transformaciones sobre los datos, tareas de clasificación, regresión, asociación y visualización. Weka está diseñado como una herramienta orientada a la extensibilidad por lo que añadir nuevas funcionalidades es una tarea sencilla. En el Anexo A se explica el funcionamiento básico de la herramienta.

3.6.3. Algoritmo J48

C4.5 es un algoritmo usado para generar árboles de decisión, surge como una extensión del algoritmo ID3. C4.5 construye árboles de decisión desde un grupo de datos de entrenamiento de la misma forma en que lo hace ID3, usando el concepto de entropía de información.

En cada nodo del árbol, C4.5 elige un atributo de los datos que más eficazmente dividen el conjunto de muestras en subconjuntos enriquecidos en una clase u otra. Su criterio es el normalizado para ganancia de información (diferencia de entropía) que resulta en la elección de un atributo para dividir los datos. El atributo con la mayor ganancia de información normalizada se elige como parámetro de decisión. *J48* es una implementación *open source* o de código abierto, en lenguaje de programación Java, del algoritmo C4.5 en la herramienta WEKA para minería de datos.

Para la determinación de la raíz del árbol en el ejemplo propuesto en la Tabla 3.2 se tienen los siguientes valores

$H(\text{resultado}/\text{tiempo})=1,373$ bit de información.

$H(\text{resultado}/\text{temperatura}) = 0,85$ bit de información.

$H(\text{resultado}/\text{humedad}) = 0,92$ bit de información.

$H(\text{resultado}/\text{viento}) = 1,41$ bit de información.

Si se analiza la situación que se plantea para los dos valores más bajos, en este caso atributo temperatura y humedad respectivamente. Al tomar temperatura como raíz queda el árbol inicial de la figura 3.4.

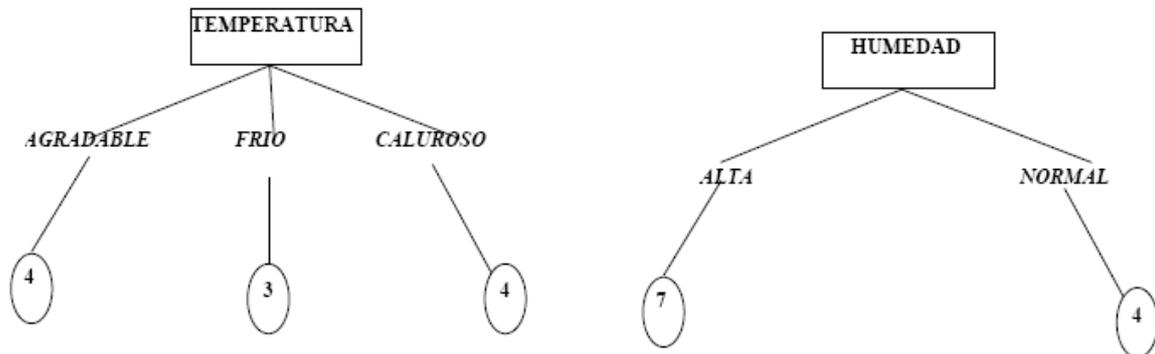


Figura 3.4. Árboles iniciales para temperatura y humedad

Cuando se trabaja con el concepto de tomar como raíz del árbol aquel atributo que menor cantidad de información *deja en el árbol* o lo que es lo mismo la que mayor Ganancia de Información posee respecto del objetivo *deporte*. La Información necesaria para identificar cualquier nodo del árbol es:

$$(3/11) \cdot \log_2(11/3) + (4/11) \cdot \log_2(11/4) + (4/11) \cdot \log_2(11/4) = 1,5726$$

En Función de ello las Ganancias de Información para cada atributo resulta:

$$GI(\text{tiempo}) = (1,5726 - 1,373) \text{ bit de información} = 0,1996$$

$$GI(\text{temperatura}) = (1,5726 - 0,85) \text{ bit de información} = 0,7226$$

$$GI(\text{humedad}) = (1,5726 - 0,92) \text{ bit de información} = 0,6526$$

$$GI(\text{viento}) = (1,5726 - 1,41) \text{ bit de información} = 0,1626$$

En este caso como raíz del árbol se tomaría el de mayor ganancia y no cambiaría nada respecto del análisis anterior. Ahora si en lugar de analizar la GI solamente se analiza la GI relativa al grado de ramificación de cada atributo resulta lo siguiente:

Para temperatura

$$(-4 \cdot \log 24 - 3 \cdot \log 23 - 4 \cdot \log 24 + 11 \cdot \log 211) / 11 = 1,5726$$

$$GIR(\text{temperatura}) = 0,7226 / 1,5726 = 0,4594$$

Para Humedad

$$(-7 \cdot \log 27 - 4 \cdot \log 24 + 11 \cdot \log 211) / 11 = 0,9456$$

$$GIR(\text{humedad}) = 0,6526 / 0,9456 = 0,6900$$

Resultando en este caso en una nueva raíz cuyo árbol final es el de la figura 3.5.

En este caso las hojas del árbol son solamente siete y las del árbol anterior eran ocho. Con este nuevo análisis se logra un árbol más óptimo que el anterior, sin perder información y eliminando la redundancia que tenía la base de datos original ya que no se tiene en consideración al atributo viento.

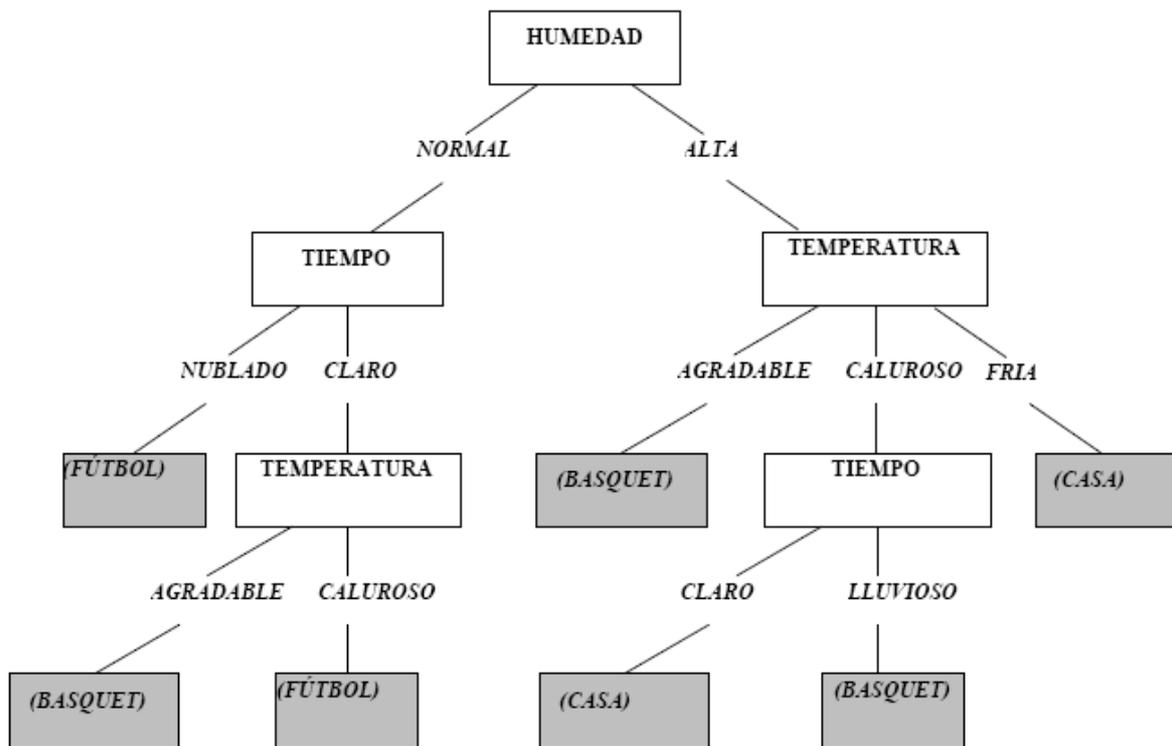


Figura 3.5 Árbol final para humedad

3.6.4. Evaluación de clasificadores

Existen diversas técnicas para la evaluación de los algoritmos clasificadores, Hernández et al (2004) describen los siguientes:

- Validación simple o *Training y test*: consisten en disponer de un conjunto de casos de memoria y un conjunto de casos de test, conjuntos diferentes, uno que entrena y otro que prueba el modelo.

- Validación Cruzada o *Cross-validation*: se divide el conjunto de casos disponibles en M subconjuntos. Se utilizará un de los subconjuntos como un conjunto de test, y los otros M-1 para hacer el entrenamiento del sistema. Se repite el proceso utilizando cada uno de los conjuntos como un conjunto de test y el resto para entrenar.

- Matriz de confusión o *Confusion Matrix*. Es una herramienta de visualización que se emplea en aprendizaje supervisado. Cada columna de la matriz representa el número de predicciones de cada clase, mientras que cada fila representa a las instancias en la clase real. Uno de los beneficios de las matrices de confusión es que facilitan ver si el sistema está confundiendo dos clases.

- Análisis de las Características de funcionamiento del receptor o Receiver Operating Characteristics, o análisis ROC por sus siglas en inglés. Representa de forma gráfica la relación entre TP, número de casos de la clase A acertados correctamente y FP número de casos de la clase A no acertados de un sistema de diagnosis en función de un parámetro de decisión.

Además, existen otras medidas de desempeño generadas con las que también se pueden evaluar los modelos como son: la precisión, la confianza y cobertura, el error cuadrático medio, la cohesión y separación.

3.7. APLICACIONES

En este capítulo se describen dos técnicas de minería de datos: reglas de asociación y árboles de decisión. La primera técnica se utiliza para extraer patrones de comportamiento entre los datos en función de la aparición conjunta de valores de dos o más atributos que suceden más frecuentemente. La segunda técnica construye un *modelo* o representación de la regularidad

existente en los datos que se pueden expresar de una manera simbólica en la forma de un conjunto de condiciones inteligibles para los seres humanos.

Una vez contruidos y validados los modelos puede usarse principalmente con dos finalidades: para que un analista recomiende acciones basándose en el modelo y en sus resultados, o bien para aplicar el modelo a diferentes conjuntos de datos. También puede incorporarse a otras aplicaciones, como por ejemplo a un sistema de análisis de créditos bancarios, que asista al empleado bancario a la hora de evaluar a los solicitantes de los créditos, o incluso automáticamente, como los filtros de spam o la detección de compras con tarjetas de crédito fraudulentas.

Tanto en el caso de una aplicación manual o automática del modelo, es necesaria su difusión, es decir que se distribuya y se comunique a los posibles usuarios.

También es importante medir lo bien que el modelo evoluciona. Aún cuando el modelo funcione bien se debe continuamente comprobar las prestaciones del mismo. Esto se debe principalmente a que los patrones pueden cambiar. Por lo tanto el modelo debe ser monitorizado, lo que significa que de tiempo en tiempo el modelo tendrá que ser re-evaluado o re-entrenado y posiblemente reconstruido completamente.

CAPÍTULO 4

Evaluación de la presencia cognitiva

CAPÍTULO 4: EVALUACIÓN DE LA PRESENCIA COGNITIVA

4.1. INTRODUCCIÓN

Las cuestiones relativas a la responsabilidad y el control en un proceso educativo son relevantes tanto en la dimensión de la enseñanza como en la del aprendizaje. Las responsabilidades del profesor son complejas en el sentido de que crean y configuran el entorno de aprendizaje. Este reto se complica aún más cuando intervienen las nuevas tecnologías. Los profesores deben crear las condiciones cognitivas y sociales adecuadas para que los estudiantes otorguen pleno sentido a su experiencia educativa.

Es precisamente la gestión del aprendizaje, tanto en sentido técnico como cognitivo, lo que está transformando de modo radical la educación superior. La gestión técnica se refiere a la organización y acceso a la información, mientras que la gestión cognitiva de la enseñanza se centra en el seguimiento cognitivo y metacognitivo de las tareas y exigencias intelectuales de los alumnos.

Los autores Garrison y Anderson (2005) expresan:

La clave para entender las prácticas educativas está en plantearse cuáles son los resultados perseguidos. En la educación superior, esos resultados se asocian con un nivel elevado de aprendizaje en que cobra especial protagonismo el aprender a pensar de modo crítico y creativo. Recientemente se han añadido otras orientaciones, como la de promover la auto disciplina, ya que no es tarea fácil conseguir que los estudiantes asuman la responsabilidad de su formación. (p. 33)

Comprobar la calidad o éxito en el proceso de aprendizaje en un entorno e-learning es una empresa más que compleja y no exenta de problemas metodológicos. El modelo de investigación práctica propuesto por Garrison y Anderson permite evaluar la presencia cognitiva, además ofrece un medio para juzgar la naturaleza y calidad de la reflexión y el discurso crítico en una comunidad de investigación que trabaja en cooperación. Estos autores presentan un conjunto de descriptores e indicadores de la presencia cognitiva que son aptos para valorar la calidad de la investigación. Se trata de emplear esos indicadores para evaluar el nivel de pensamiento y discurso críticos en el gran número de mensajes que se generan durante las sesiones de e-learning.

Con el objeto de enriquecer el marco teórico se indaga en la aplicación de los indicadores de presencia cognitiva en otras investigaciones. A continuación se describen los trabajos seleccionados.

4.2. IDA FALLAS MONGE

En el trabajo de Ida Fallas Monge (2008) denominado *¿Está mi foro produciendo pensamiento crítico?*, se describe un estudio sistemático de análisis de contenido en un foro de discusión en el que se utiliza el modelo de Garrison, Anderson y Archer (2001).

Para realizar el estudio del contenido selecciona uno de los diez foros de un curso a nivel de doctorado, trabajado durante ocho días, en el mes de marzo del 2008 y que acumula un total de 60 intervenciones por parte de los 15 estudiantes del grupo, ante una propuesta de seis preguntas por parte de la profesora. Las respuestas son distribuidas y cuantificadas en cada una de las cuatro categorías que el modelo de investigación práctica establece para determinar la presencia cognitiva en un foro de discusión: hecho desencadenante, exploración de ideas, integración y resolución, ver tabla 4.1.

Al realizar el análisis de las intervenciones excluye 15 participaciones correspondientes a la primera intervención de cada estudiante, ya que en ellas se da respuesta a las seis preguntas planteadas en el foro, las cuales demandan un aporte que en realidad corresponde a cada una de las fases del modelo de investigación práctica. Por lo tanto, clasifica el resto de los 45 mensajes de los alumnos.

En los resultados de la clasificación detecta que la fase de exploración de ideas es la de mayor frecuencia en el foro, con un 44% de los mensajes en esta categoría. Seguidamente se ubican 12 mensajes en el primer nivel de exploración, que representan el 27% de los mensajes. En la categoría de integración identifica 9 mensajes que constituyen un 20% de la interacción; finalmente, el nivel superior es el de menor número de participaciones con un total de cuatro aportes, que representan el 9 % de la discusión.

Para finalizar el estudio analiza los mensajes de cada estudiante de acuerdo con la categoría respectiva, excluyendo el primer mensaje ya que calza en todas las fases.

Tabla 4.1 *Fases del Modelo de Investigación Práctica*

Fase	Descriptor	Indicador	Se asocia con
Hecho Desencadenante	Evocativo (inductivo)	Reconocer el problema	Conceptualización del problema
Exploración de ideas	Inquisitivo (divergente)	Divergencia Intercambio de Información Sugerencias Lluvia de Ideas Saltos intuitivos	Búsqueda de ideas e información relevante
Integración	Tentativo (convergente)	Convergencia Síntesis Soluciones	Elaborar una solución o explicación apropiada
Resolución	Comprometido (deductivo)	Aplicar Comprobar Defender	Evaluación de la viabilidad de la solución propuesta

Fuente: Fallas Monge (2008, p. 5)

La tabla 4.2 muestra que los estudiantes que alcanzan la fase de resolución, también tienen mensajes en la fase de integración, y además no tienen ningún mensaje en la fase inicial correspondiente a la categoría de hecho desencadenante.

Tabla 4.2. *Clasificación de los mensajes de cada estudiante*

Estudiante	Fase 1	Fase 2	Fase 3	Fase 4
N° 1	2	3	0	0
N° 2	0	0	0	0
N° 3	1	1	1	0
N° 4	0	1	1	0
N° 5	0	1	2	1
N° 6	2	2	0	0
N° 7	0	3	1	0
N° 8	5	4	1	0
N° 9	0	1	0	0
N° 10	0	1	1	2
N° 11	0	2	1	0
N° 12	0	0	1	10
N° 13	1	1	0	0
N° 14	0	0	0	0
N° 15	1	0	0	0
TOTAL	12	20	9	3

Fuente: Fallas Monge (2008, p. 10)

El análisis del foro de acuerdo con el modelo de investigación práctica de Garrison et al. (2001), da resultados que confirman lo establecido por los autores, quienes indican que la tendencia general en las discusiones en línea es la de resolver muy bien las dos primeras fases, la tercera no tanto y la cuarta dejarla al margen. En el foro estudiado se aprecia que una forma de lograr alcanzar el cuarto nivel, es incluir en la propuesta inicial una pregunta que enfatice precisamente el nivel de aplicación o resolución.

De acuerdo con los resultados obtenidos, se nota que un grupo de discusión tiende a concentrarse en aspectos de la segunda fase, o etapa de exploración, en los cuales se busca entender el problema y buscar la información relevante. Al respecto Garrison y Anderson (2005), indican que el reto educativo está en gestionar y orientar esa fase hacia el pensamiento divergente de modo que sirva de transición para la fase siguiente.

4.3. RENÉ ARROYO ÁVILA, GERARDO REYES LÓPEZ Y MYRNA GARCÍA BENCOMO

Los autores mexicanos, René Arroyo Ávila, Gerardo Reyes López y Myrna García Bencomo (2010) realizan una investigación con el objetivo general de caracterizar la presencia docente en las discusiones virtuales y su relación con el desarrollo de las habilidades cognitivas (presencia cognitiva), en los estudiantes en la modalidad en línea en la Universidad Autónoma de Chihuahua. De este estudio se extrae sólo lo referente a presencia cognitiva para realizar una comparación con los otros trabajos investigados.

Los participantes son 21 estudiantes, que cursan sus estudios profesionales en la modalidad en línea de la Facultad de Contaduría y Administración de la Universidad Autónoma de Chihuahua, en el semestre enero a julio de 2009.

El foro de discusión que estudian se enmarca en un curso de *Administración de la Mercadotecnia*. El tema que se trata en el foro de discusión es la demografía en los mercados, que corresponde al objeto de estudio tres: identificación y selección de mercados. Los estudiantes previamente trabajan los objetos de estudio uno y dos: la naturaleza y el panorama del marketing.

El objetivo del profesor con el foro es construir conjuntamente con los alumnos, el conocimiento sobre la forma en la que los cambios demográficos influyen en los mercados de bienes o servicios.

Los investigadores analizan el texto de cada uno de los mensajes de los estudiantes en el foro, con lo que cualitativamente determinan cuales mensajes expresan confusión con el contenido del tema, cuales tienen claridad con el tema que se está tratando y por consiguiente han logrado construir el significado de su aprendizaje. También detectan cuales mensajes expresan divergencia de ideas con lo que se plantea en el foro, cuales mensajes expresan sugerencias sobre el tema y cuales ofrecen intercambio de información al resto de sus compañeros. Identifican cuales mensajes generan síntesis del contenido que se está construyendo, así como la expresión de convergencia con las ideas de sus compañeros. Finalmente establecen cuales mensajes expresan defensa de sus ideas, cuales asienten una comprobación de sus ideas y cuales aplican nuevas ideas a la discusión. Las clasificaciones que dan a los mensajes no son excluyentes, es decir que un mensaje puede ser clasificado dentro de varios indicadores en cada una de las categorías analizadas dentro del componente presencia cognitiva, del modelo de comunidad de aprendizaje.

En la categoría *hecho desencadenante*, que es donde el estudiante ubica el problema de estudio, detectan 13 mensajes en los que los estudiantes reconocen la claridad del problema planteado en el foro y expresan buen conocimiento inicial del tema; además de tres mensajes que expresan confusión al iniciar el debate y que necesitan que el profesor los centre en el debate y los incentive para construir adecuadamente su aprendizaje.

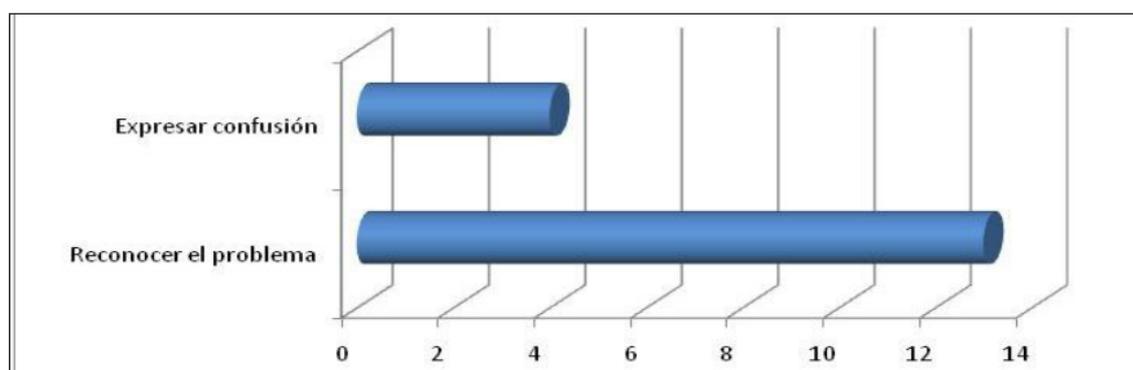


Figura 4.1 Categoría *hecho desencadenante*, dentro del componente *presencia cognitiva*
Fuente: Arroyo et al. (2010, p. 16)

En 25 mensajes los estudiantes intercambian información con sus compañeros y con el profesor. Las aportaciones que progresivamente se hacen en el foro denotan mayor conocimiento de lo que se expresa. Además, quienes inicialmente muestran confusión, al avanzar las discusiones van expresando con mayor claridad los conceptos que se espera formen parte de su conocimiento logrado.

En la categoría *exploración*, encuentran 11 aportaciones donde los estudiantes expresan sugerencias a sus compañeros sobre el tema y cuatro mensajes en los que expresan divergencia de ideas con lo planteado por sus compañeros. Todo esto refuerza la construcción del contenido y del aprendizaje que se busca con el foro de discusión.

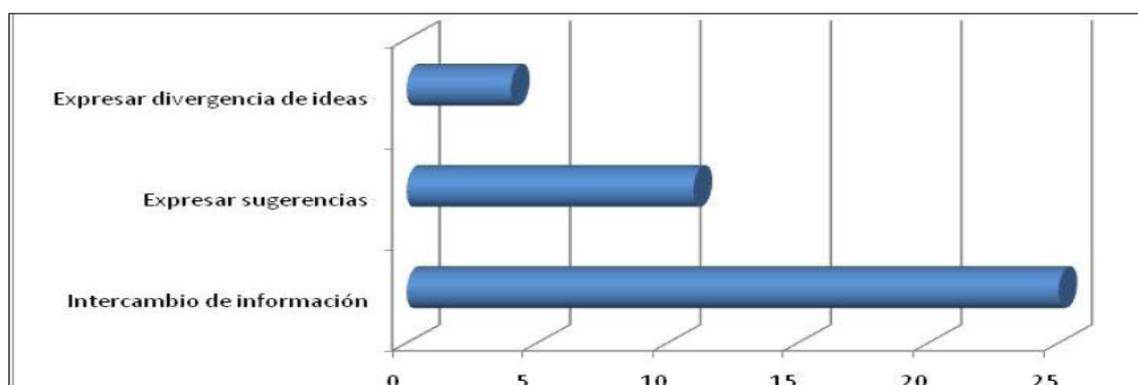


Figura 4.2 Categoría *exploración*, dentro del componente *presencia cognitiva*

Fuente: Arroyo et al. (2010, p. 17)

En la categoría *integración*, observan que en 15 mensajes al final de sus aportaciones los alumnos a manera de conclusión de lo aprendido, hacen una breve síntesis del tema tratado y tres mensajes donde expresan la convergencia de ideas entre los propios estudiantes.

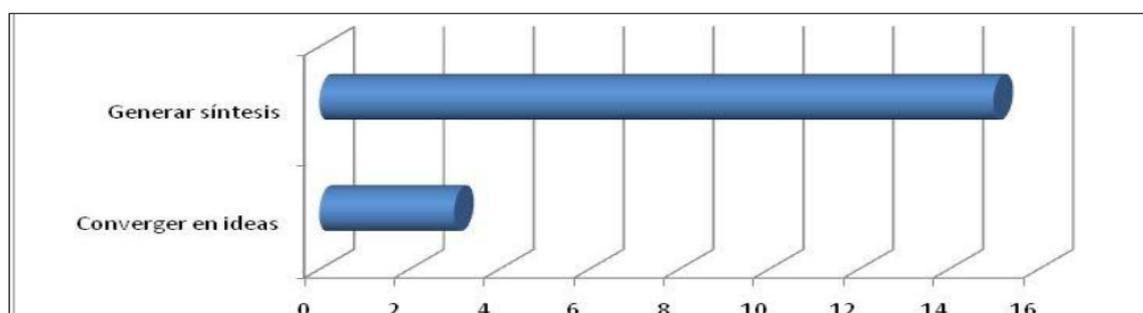


Figura 4.3. Categoría *integración*, dentro del componente *presencia cognitiva*

Fuente: Arroyo et al. (2010, p. 17)

En la categoría *resolución*, detectan 14 mensajes donde se aprecia que los estudiantes comprueban sus ideas con los demás, cuatro aportaciones son para defender sus ideas, sobre todo cuando otro estudiante cuestiona el contenido de su aporte y uno más que expone nuevas ideas sobre el tema planteado.

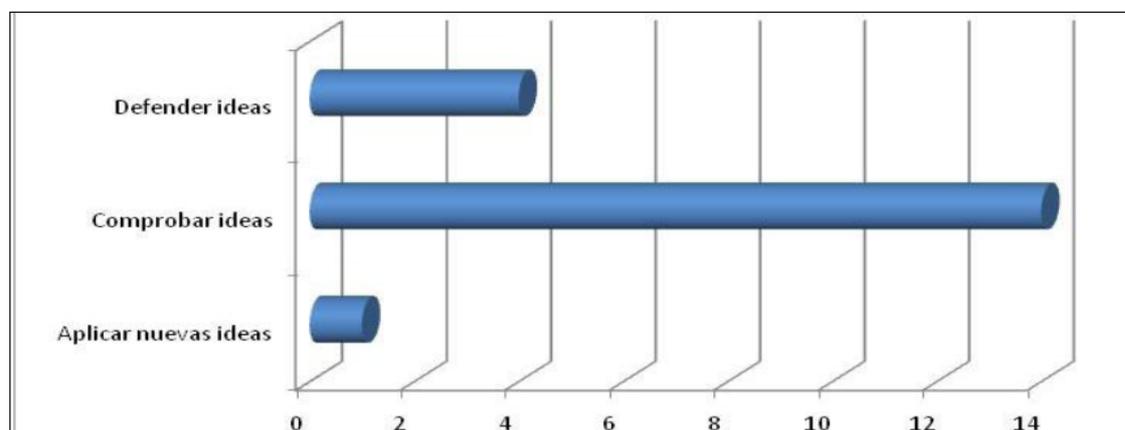


Figura 4.4 Categoría *resolución*, dentro del componente *presencia cognitiva*
Fuente: Arroyo et al. (2010, p. 18)

Con lo observado en este caso de estudio, se puede determinar que la construcción de conocimiento y el desarrollo de habilidades cognitivas en el estudiante (presencia cognitiva), en un ambiente de aprendizaje en línea, se puede lograr a través de una adecuada participación e interacción del profesor (presencia docente), sobre todo en actividades colaborativas donde la interacción facilite compartir ideas y experiencias, que conducidas adecuadamente permita que se cumpla el objetivo de aprendizaje de una manera más pronta y expedita.

4.4. MAGDALENA CATALINA BUSSONE

Magdalena Catalina Bussone (2006) es profesora de la Facultad de Lenguas de la Universidad Nacional de Córdoba y autora de la ponencia *Una experiencia de comunidad de indagación: hacia la comprensión de las interacciones realizadas en una comunidad virtual de aprendizaje*.

En este trabajo realiza un análisis de algunas de las interacciones que surgen en un ambiente de aprendizaje virtual creado para la formación de docentes de Italiano en el cual la autora del mismo participa de manera directa como co-tutora de formación on-line.

Aplica el modelo de comunidad de indagación de Garrison, Anderson y Archer (2000) para analizar las interacciones virtuales colaborativas de una comunidad de aprendizaje blended learning. De manera particular, muestra cómo se realizan los tres constructos propuestos por el modelo aplicado: la presencia social, la presencia cognitiva y la presencia docente. El estilo horizontal y participativo que caracteriza las comunidades de aprendizaje es ilustrado, a través de la aplicación del modelo elaborado por el equipo canadiense, para señalar de qué manera se construyen las diferentes presencias en el grupo.

Los registros son tomados de un foro que forma parte de las actividades académicas desarrolladas en el curso de formación en red para docentes de lengua italiana como lengua extranjera, con participantes de distintos contextos geográficos.

Los ejemplos pertenecen a un foro del Módulo titulado *Didáctica del Italiano en Perspectiva Intercultural* que se dicta en la modalidad *blended learning*, de la Universidad Ca' Foscari de Venecia, Italia. La duración de las actividades online del citado módulo es de un mes, cada semana de trabajo implica una nueva actividad para realizar en el foro y prevé un mínimo de cuatro intervenciones por cada participante.

Propone una metodología cualitativa a través del análisis del discurso, la unidad de análisis seleccionada son unidades sintácticas como la palabra, la proposición, la frase o el párrafo.

Se extraen dos de los once ejemplos que describe el trabajo, con el objetivo de mostrar el detalle del análisis que realiza:

Autore	Messaggio
<p data-bbox="199 1608 295 1637">Paola_N</p> <p data-bbox="199 1727 391 1778">Registrato: 13 Giu 2005</p> <p data-bbox="199 1787 359 1816">Messaggi: 185</p>	<p data-bbox="442 1608 1166 1659">☐ Oggetto: Re: Riflessione sulla cronemica Inviato: 16 Set 2006 - 00:09</p> <p data-bbox="1273 1630 1437 1659"> riporta </p> <p data-bbox="442 1697 1396 1816">Intervengo nel dibattito sulla cronemica, dove la mia esperienza innanzi tutto non può che confermare lo stereotipo dell'italiano poco puntuale (...) Credo che ogni cultura abbia una "cronemica dello sguardo", e che ogni deviazione rispetto a questo implicito orologio interno possa dare adito a fraintendimenti. Che ne pensate? Paola</p>

Figura 4.5. *Ejemplo 1*

Fuente: Bussone (2006, p. 6)

Identifica *Presencia cognitiva* que se enmarca en la categoría *exploración* a través del *intercambio de información*: “creo que cada cultura tiene una cronémica de la mirada...”.

La *Presencia social*, enmarcada en la categoría *comunicación cohesiva*, se refleja en la *formulación de preguntas* que promueve el *intercambio* y la *colaboración*. El estudiante finaliza su mensaje diciendo: “¿qué piensan ustedes?”.

Autore	Messaggio
Chiara_S	<p>☐ Oggetto: Re: Didattica dell'italiano: seconda attività</p> <p>11/09/06-17/09 Inviato: 16 Set 2006 - 13:06 </p> <hr/> <p>Cara Rosalia,</p> <p>Registrato: 14 Giu 2005</p> <p>Messaggi: 227</p> <p>cito a conferma di quanto tu dici un episodio che all'inizio mi ha lasciato diciamo perplesso. In Turchia (...)  Non ti dico la mia sorpresa, e quando ho chiesto (...) Inutile dire (...) alla fine anche io mi sono abituata (che tristezza pero') Chiara</p>

Figura 4.6. *Ejemplo 2*

Fuente: Bussone (2006, p. 6)

Detecta *Presencia social* enmarcada en *comunicación interactiva*: el alumno retoma/continúa un mensaje de otro compañero: “cito para confirmar lo que vos decís” y *comunicación afectiva*, se expresan emociones: “...me dejó perpleja”; “Qué tristeza”; además se recurre al empleo de *emoticons*.

La *Presencia cognitiva* se encuadra en la *exploración*: hay un intercambio de información cuando se formula: “En Turquía...”.

En la conclusión Bussone (2006) expresa que:

Como se ha dicho anteriormente, en las *comunidades de indagación* prevalece el aspecto constructivo y social de la formación: la interacción *vertical* tutor-alumno y la comunicación interactiva *horizontal* alumno-alumno son algunos de los aspectos-clave. Paradojalmente el estar sólo frente a la PC se transforma en una condición para realizar el aprendizaje colaborativo virtual. No obstante, desde el punto de vista psicológico y *relacional*, las interacciones constituyen un elemento fundamental para el aprendizaje. Se impone una toma de conciencia por parte de los docentes en el uso de las TICs y en el conocimiento de la complejidad de las interacciones que se realizan en las comunidades de aprendizaje. (p. 12)

Además a modo de síntesis presenta la tabla 4.3.

Tabla 4.3 *Síntesis Presencias*

Elementos	Categorías	Indicadores (ejemplos)
Presencia Social	Comunicación Afectiva	Expresa emociones, uso de emoticons.
	Comunicación Abierta	Continúa un mensaje, lo relaciona con otros.
	Comunicación Cohesiva	Habla de un “nosotros” estimula la colaboración.
Presencia Docente	Gestión de la Organización	Programa, calendario, define tópicos de discusión.
	Facilitación del discurso	Mantiene la comunidad de estudio.
	Enseñanza directa	Centra el debate, envía feedback y evaluación.
Presencia Cognitiva	Evento desencadenante	Sentimiento de confusión.
	Exploración	Intercambio de Información.
	Integración	Conexión de ideas.
	Resolución	Aplicación de nuevas ideas.

Fuente: Bussone (2006, p. 12)

4.5. MARÍA ESTELA RIPA

En su trabajo *¿Cómo abordar el estudio de una comunidad de aprendizaje blended learning?* María Estela Ripa (2007) estudia las interacciones virtuales colaborativas de una comunidad de aprendizaje blended learning, aplicando el modelo de comunidad de indagación de Garrison, Anderson, Rourke, Garrison y Archer (2001) y la metodología del análisis de contenido cuantitativo. Muestra la utilidad del modelo para el análisis y caracterización de las interacciones colaborativas virtuales, fundamentalmente la flexibilidad de los constructos presencia social y presencia docente para adaptarse a diversas configuraciones didácticas. El trabajo aporta además enfoques novedosos: Utiliza el modelo para analizar cómo se construyen las distintas presencias en el grupo, lo cual permite caracterizar el estilo participativo y horizontal de la comunidad de aprendizaje. Propone una nueva categoría (presencia vicaria) para aplicar el modelo a la modalidad blended learning y se señala la falta de indicadores para la actividad de auto-organización grupal que generalmente implican las tareas colaborativas.

La autora acota el estudio al análisis de las interacciones asíncronas del foro del grupo que muestra mayor actividad en la cohorte 2004 de la Escuela de Jóvenes. Este foro pertenece a una comunidad de aprendizaje compuesta por 10 alumnos y un tutor. Ninguno de ellos tiene experiencia previa en cursos a distancia o con componente virtual. El foro funciona durante 7

meses, acompañando el desarrollo de 8 módulos temáticos. El estudio se realiza sobre las interacciones de los dos meses iniciales.

Utiliza las categorías e indicadores desarrollados para el modelo de comunidad de indagación (Garrison, Anderson & Archer, 2000; Rourke et al., 1999; Garrison et al., 2001; Anderson et al. 2001) para el análisis del contenido de los mensajes. El objetivo es explorar la utilidad del modelo para evaluar la actividad del foro virtual y específicamente el componente colaborativo que hace a la comunidad de aprendizaje.

La unidad de análisis escogida fue variando a lo largo del estudio. Finalmente, luego de varias pruebas adopta como unidad el mensaje. Viendo que en un mismo mensaje hay elementos de distintas categorías y puede ser forzado asignarlo sólo a una, decide adoptar una metodología usada en varios estudios que aplican este modelo. Toma como unidad el mensaje y para cada uno se realiza una serie de decisiones binarias: por ejemplo, para evaluar presencia docente, toma para cada mensaje tres decisiones: el mensaje contiene/no contiene diseño instruccional; contiene/no contiene facilitación del discurso; contiene/no contiene instrucción directa. Asimismo identifica los mensajes que contienen los tres tipos de categorías y los que no tienen ninguna (Rourke & Anderson, 2001).

Esta metodología de codificación resulta sencilla, práctica y simple. Analiza 27 mensajes, compuestos por 203 frases y distribuidos en 8 conversaciones. Casi la mitad de los mensajes 48%, contiene los tres elementos del modelo, y la otra mitad 48% tiene al menos dos elementos. Solo el 4% de los mensajes contiene un solo elemento. Los elementos que aparecen con mayor frecuencia son presencia social y presencia docente, en el 100% y 96% de los mensajes respectivamente, mientras que presencia cognitiva aparece en el 48%.

En presencia social, el 85 % de los mensajes contiene simultáneamente indicadores de dos o tres categorías de este elemento. Entre ellas destaca la *respuesta cohesiva*, presente en todos los mensajes 100%, y la *respuesta interactiva*, presente en el 78% de los mismos.

En presencia docente, observa que el 48% de los mensajes contiene entre 2 y 3 categorías, mientras que prácticamente la otra mitad 48% contiene indicadores de 1 sola categoría. Las

categorías más frecuentes son *diseño instruccional* 67% y *facilitación del discurso* 56%. *Instrucción directa*, en cambio, está presente en el 37% de los mensajes.

En cuanto a presencia cognitiva, según la autora, la clasificación presenta más dificultades de interpretación, con lo cual los resultados son más relativos. De todos modos observa que el 52% de los mensajes no contiene elementos de presencia cognitiva. Dentro del 48% que sí contiene, encuentra el 33 % que contiene indicadores de dos o más categorías, siendo *exploración*, la categoría con mayor frecuencia 41%, luego *elemento disparador* 37% e *integración* 26%. La categoría *solución* solo verifica en el 4%.

Tabla 4.4 Frecuencia y porcentaje de mensajes por categorías de presencia

Presencia social	F	%	Presencia docente	F	%	Presencia cognitiva	F	%
Cohesiva	27	100%	Diseño instruccional	18	67%	Exploración	11	41%
Interactiva	21	78%	Facilitación del discurso	15	56%	Evento disparador	10	37%
Afectiva	18	67%	Instrucción directa	10	37%	Integración	7	26%
						Solución	1	4%

Fuente: Ripa (2007, p. 210).

La Presencia Social es el elemento más fácil de codificar, en el sentido que menos dudas presenta a la hora de clasificar los mensajes. Sin embargo surgen de la aplicación a este estudio algunas interesantes observaciones:

La categoría *interactiva*, presente en el 78% de los mensajes, incluye algunos indicadores que dependen mucho del tipo de software que utilizan y del entrenamiento que tengan o las instrucciones que reciban los participantes en el uso de la herramienta de comunicación.

Presenta el siguiente ejemplo, uno de los indicadores es ‘continuar una conversación’ y se define como usar la función ‘responder’ del software. En el caso estudiado se da una instrucción explícita para la estructuración del foro de grupo: abrir sólo una conversación (tópico) por actividad. Por tanto el primero en participar abre el tópico y los demás deben usar la función ‘responder’ para hacer sus aportes, de modo que los mensajes queden. En el caso de que el software utilizado no ofrezca esa funcionalidad ‘responder’ es aconsejable desvincular el indicador del software y relacionarlo con la cita, parafraseo o referencia al mensaje de otro.

El indicador ‘hacer preguntas’ no contempla acciones -claramente vinculadas a la interacción- como la formulación de pedidos y sugerencias. Ripa (2007) presenta los siguientes ejemplos tomados del foro:

- *Ya que no nos pusimos de acuerdo en XX... Quizás podamos tomar la actividad 2 de análisis de artículos periodísticos.*
- *Hola chicos, acá les envío la actividad número 3, (...) contesten algo, así trato de modificarlo o interactuamos un poco más.* (p. 211)

Estas interpelaciones muestran que el alumno sabe que no está solo, que hay interlocutores a quienes se puede dirigir y en este sentido reflejan un aspecto central de la presencia social.

Del mismo modo, destaca que no aparece en el modelo un indicador que recoja los pedidos de disculpas y los agradecimientos, ni tampoco un elemento característico de la comunidad estudiada: la comunicación de desvíos o logros frente a lo acordado o, simplemente, el estado de avance. Ripa (2007) muestra algunos ejemplos:

- *Después de todo este tiempo... me pongo a escribir... disculpen que no haya estado.*
- *Bueno, tarde pero seguro, les mando lo que fui haciendo (me falta la c).*
- *Hola chicos, quería comentarles lo que he hecho (...). En la reunión les daré detalles.*
- *Disculpen la ausencia, rendí y todo salió bien. Ya estoy de vuelta.*
- *Buenas a todos, disculpen mi desaparición pero tuve que ir a Rosario a rendir una materia. Me fue bien así que me van a volver a disculpar ya que este viernes voy a fal-tar nuevamente (...) Sin embargo estoy haciendo las tareas y N. me comenta los en-cuentros, así que no estoy descolgada. (...)*
- *Gente: gracias por el esfuerzo en que llegue con el horario, pero me di cuenta que por más magia que haga no me será posible ir.* (p. 211)

Estos contenidos, presentes en el material analizado, expresan claramente la esencia de la categoría presencia social: son intercambios recíprocos caracterizados por una mutua conciencia y el reconocimiento del otro, evidencias de que hay otro que está esperando y que se lo tiene en cuenta (Garrison, Anderson and Archer, 2000, p. 100).

El análisis de la Presencia Cognitiva lo realiza tomando como referencia el estudio de Mayer (2004) sobre diversos marcos de análisis para las discusiones on line. Allí se aplica el modelo de presencia cognitiva encontrando porcentajes no muy diferentes a los obtenidos en este trabajo. El 18% de los mensajes analizados por Mayer entró en la categoría de evento

disparador; el 59% en exploración e integración y el 20% en solución. Los resultados de su estudio arrojan 37%, 67% y 4% respectivamente.

Sin embargo, destaca que presencia cognitiva es el elemento del modelo más difícil de aplicar en la codificación de los mensajes, con lo cual estos porcentajes deben ser revisados a la luz de un ajuste de los criterios utilizados.

Las dificultades encontradas las sintetiza en que básicamente la consigna del foro de grupo no dirige la tarea hacia la aplicación de las categorías del constructo presencia cognitiva del modelo.

Ripa (2007) indica que la propuesta para el foro de grupo básicamente trata de poner en común la respuesta personal a la consigna (generalmente análisis de un texto) y eventualmente intercambiar comentarios sobre estas con los compañeros. Esta tarea poco tiene que ver con discutir o debatir ideas, argumentar, negociar y solucionar problemas, que es el modo en que se entiende la construcción conjunta del conocimiento que asume el modelo.

Los resultados que obtiene son consistentes con la observación de Meyer (2004) acerca de que la naturaleza de la consigna (evento disparador) influye en el tipo y nivel de respuesta de los alumnos. Y que el modelo de análisis de la interacción debe corresponderse con el tipo de objetivos y tareas definidas en el diseño. En este sentido, expresa que el modelo de pensamiento crítico que subyace a la definición de presencia cognitiva no es el más adecuado para recoger la actividad de este foro.

El intento de aplicar el modelo le permitió sin embargo identificar una característica distintiva relacionada con la modalidad blended learning del curso analizado: la referencia a la presencia cognitiva del encuentro presencial, es decir una presencia cognitiva virtual *vicaria* o indicativa de la construida en el encuentro presencial.

4.6. APORTES A LA INVESTIGACIÓN

Los trabajos que se presentan en este capítulo permiten analizar la aplicación del modelo de Garrison y Anderson (2005) para abordar el estudio de las contribuciones que las discusiones en línea puedan hacer en la promoción de estrategias de pensamiento crítico.

El aporte del trabajo de Ida Fallas Monge (2008) es que en su análisis detecta posibles relaciones de co-ocurrencia de los indicadores de presencia cognitiva, es decir posibles reglas de asociación. En su conclusión expresa:

Los estudiantes que alcanzaron la fase de resolución, también tienen mensajes en la fase de integración, y además no tienen ningún mensaje en la fase inicial correspondiente a la categoría de hecho desencadenante. (p. 9)

El aporte de Arroyo et al. (2010) es la descripción detallada de la heurística de categorización, por ejemplo, describe:

Se analizó el texto de cada uno de los mensajes de los estudiantes en el foro, con lo que cualitativamente se determinó cuáles mensajes expresaban confusión con el contenido del tema, cuáles tenían claridad con el tema que se estaba tratando y por consiguiente habían logrado construir el significado de su aprendizaje. (p. 13)

El aporte de Magdalena Catalina Bussone (2006) es que muestra el mensaje original y la forma en que se realiza el análisis. En cada caso ejemplifica la aplicación del marco teórico en un contexto real y las dificultades de clasificar cada frase en alguna de las categorías propuestas.

Las conclusiones de María Estela Ripa (2007) con respecto a la presencia cognitiva son muy valiosas. Ella expresa:

La presencia cognitiva ha sido el elemento del modelo más difícil de aplicar en la codificación de los mensajes. Las dificultades encontradas pueden sintetizarse en que básicamente la consigna del foro de grupo no dirige la tarea hacia la aplicación de las categorías del constructo presencia cognitiva del modelo. (p. 214)

Aunque queda mucho trabajo por hacer para perfeccionar esta herramienta, el modelo de investigación práctica, con sus indicadores, es un buen método heurístico para guiar y evaluar la naturaleza y calidad de la presencia cognitiva.

CAPÍTULO 5

Evaluación del curso de e-learning

CAPÍTULO 5: EXTRACCIÓN DE CONOCIMIENTO

5.1. INTRODUCCIÓN

En función del objetivo general, extraer conocimiento de los foros de los cursos de e-learning con el objeto de apoyar la mejora continua del proceso de enseñanza-aprendizaje, se propone un estudio de caso. Este se basa en el análisis de los mensajes enviados a los foros de discusión de dos cursos de e-learning desarrollados en la Universidad Nacional de San Juan.

El primero de los cursos denominado *Enseñar y Aprender en el Aula Virtual* se desarrolla durante el periodo agosto-diciembre de 2008 y el segundo *Prácticas Educativas en Entornos Virtuales* en el periodo abril- junio de 2011. Estos cursos corresponden a actividades de posgrado dirigidas a docentes de la Universidad, implementados en la plataforma moodle y organizados por el mismo grupo de profesionales que forman parte de los proyectos de investigación mencionados en el primer capítulo.

En los dos cursos objeto de este estudio, las actividades de aprendizaje grupales de cada unidad se implementan en foros. Cada grupo está compuesto por cuatro o cinco estudiantes y un tutor que conforman una comunidad de investigación.

Los tres elementos básicos de una comunidad de investigación que deben ser tenidos en cuenta al planificar y estructurar una experiencia de e-learning son: la presencia cognitiva, la presencia social y la presencia docente.

Los investigadores de los proyectos cumplen el rol de tutores de los grupos. Ellos abordan el estudio de los foros para la evaluación de las distintas presencias, relevan datos y generan informes con sus resultados.

La propuesta inicial de este trabajo es tomar todos los resultados de las evaluaciones y analizarlos con las técnicas de minería de datos. El problema que surge es que la forma de recolección, codificación y organización de la información es heterogénea y los datos no están completos o no están adaptados a las técnicas a aplicar.

Es por este motivo que se genera una heurística e instrumentos para normalizar las tareas de recolección, codificación y organización que se describen en el Anexo B. A partir de esta normalización se procesan los mensajes de los foros para obtener los datos a analizar con las técnicas de minería de datos. Esta tarea es muy minuciosa, requiere mucho tiempo y sobre los mismos textos deben intervenir tres analistas, para realizar los consensos y obtener la codificación final.

Se decide trabajar sobre los foros de tres grupos y sólo con presencia cognitiva. Es por eso que la cantidad de datos a analizar está muy limitada y el estudio solo plantea un marco de trabajo que organiza el proceso de análisis de la presencia cognitiva y que puede ser extendido a la presencia social y a la presencia docente.

En este capítulo se describe el “caso” que se estudia, el proceso realizado para la categorización de los mensajes de los foros de aprendizaje de los cursos de e-learning y la aplicación de dos técnicas de minería de datos para extraer conocimiento que contribuya a la *mejora del proceso de enseñanza-aprendizaje*

5.2. **CURSO ENSEÑAR Y APRENDER EN EL AULA VIRTUAL**

El diseño, planificación e implementación del Curso estuvo a cargo de: Waisman, E.; Olivares, M.; Gómez, C. y Font, G. Se transcribe a continuación parte de la propuesta académica.

5.2.1. Presentación y Objetivos

Los EVA son espacios cuyas características permiten facilitar la formación más allá de los límites espacio-temporales que condicionan la enseñanza presencial y que, por su flexibilidad y versatilidad, posibilitan la adecuación a distintos requerimientos de educación.

En el contexto actual en que predominan ofertas sobre la *mercantilización de la educación*, se hace necesario ocupar genuinos y valiosos espacios de formación no-presencial y semipresencial. La universidad pública tiene una especial responsabilidad para atender las crecientes demandas educativas de una población con intereses, expectativas y capacidades

diversas, lo que implica la formulación de nuevos conceptos y estrategias formativas innovadoras, que vayan más allá de los enfoques convencionales.

Objetivos centrales:

- Explicitar las particularidades, posibilidades y desafíos de la Formación Docente No-Presencial en entornos virtuales de aprendizaje (EVA), como innovación en la formación docente de grado/postgrado.
- Posibilitar la apropiación de estrategias pedagógico-didácticas que permitan construir propuestas académicas en EVA, en la modalidad semi-presencial y no-presencial

The screenshot shows the interface of a virtual classroom. At the top, the course title "Curso de Postgrado 'Enseñar y Aprender en el Aula Virtual'" is displayed in green. Below the title, there is a navigation menu on the left with categories like "Personas", "Actividades", "Buscar en los foros", and "Administración". The main content area features a "Diagrama de temas" with a globe image and a message in orange text: "Estamos transitando los últimos días del Curso. Por favor completen el cuestionario 'Su opinión nos interesa...' y vayan terminado las actividades pendientes... El Aula está próxima a cerrarse. 😊". Below this, there are sections for "Información General" and "Espacios de Interacción". On the right side, there is a "Novedades" section with a list of recent activity updates, including dates and user names like "mario olivares" and "Elena Rut Waisman".

Figura 5.1 *Curso de Postgrado Enseñar y Aprender en el Aula Virtual*

Fuente: <http://www.aulasiglo21.com.ar/campus2>

5.2.2. Contenidos y metodología

Unidad 1: Entornos Virtuales de Aprendizaje.

Unidad 2: Aprendiendo en el aula virtual.

Unidad 3: Enseñando en el aula virtual.

El curso tiene un abordaje que potencia la reflexión a partir de lecturas y de la revisión de las propias prácticas como docentes, con el propósito de analizar, debatir y construir una propuesta didáctica micro (web-quest), como expresión superadora del proceso inicial de aprendizaje.

Las actividades de aprendizaje se realizan de manera individual y/o grupal según se indique en las consignas. Con el fin de garantizar el debate y el intercambio se implementan foros con la coordinación del tutor/a, referidos a las temáticas de las tres unidades.

5.3. **CURSO PRÁCTICAS EDUCATIVAS EN ENTORNOS VIRTUALES**

En los siguientes apartados se esboza en general la Propuesta Académica del Curso, cuya autoría coincide con la del anteriormente descrito.

5.3.1. Presentación y Objetivos

El curso *Prácticas educativas en Entornos Virtuales* tiene la siguiente presentación:

La educación ha sufrido grandes cambios, tanto en la forma de circulación y apropiación del conocimiento, como en el surgimiento de nuevos espacios y formas de intervención docente, a partir del desarrollo tecnológico. Sabemos también que nos enfrentamos a un cambio paradigmático, que exige a las distintas instituciones educativas y sus profesores repensarse a sí mismos... en consecuencia, se plantean nuevos desafíos que nos exigen del mismo modo, repensar las prácticas de enseñanza y los procesos de aprendizaje que se desarrollan en su interior. Por lo que la incidencia de las tecnologías digitales en el campo de la educación, nos está indicando el surgimiento de un nuevo paradigma pedagógico-tecnológico y comunicacional.

Como podemos observar, es un tema que nos confronta con el abordaje de un espacio de complejidad, ya que permite analizar y reflexionar sobre la interrelación de la tarea educativa con sus actores, la institución, el contexto y la implementación significativa y pertinente de la tecnología, lo que implica la formulación de nuevos conceptos y estrategias formativas innovadoras que vayan más allá de los enfoques convencionales (Waisman, E; Gómez, C; Olivares, M; Font, G, 2011, p.1).

Los objetivos a alcanzar en este curso:

- Compartir ideas y propuestas para conocer prácticas pedagógicas actuales, desafíos y posibilidades, trabajando en entornos virtuales de enseñanza aprendizaje (EVA).

- Propender a que las razones por las cuales los profesores deseen diseñar aspectos de sus cátedras en EVA, sean las de generar prácticas innovadoras e inclusivas socialmente.
- Potenciar los procesos de enseñanza con la utilización de entornos virtuales para favorecer el aprendizaje significativo y colaborativo de los estudiantes.

Figura 5.2- Curso *Prácticas Educativas en Entornos Virtuales*

Fuente: <http://www.ecampus.unsj.edu.ar/course/view.php?id=106>

5.3.2 Contenidos y metodología

Unidad 0: Presentándonos. (Figura 5.1.)

Unidad 1: La Enseñanza en Entornos Virtuales.

Unidad 2: Recursos Educativos en la Web

El curso tiene un abordaje que potencia la reflexión de los límites y posibilidades de la construcción del conocimiento en la red, atendiendo a la dimensión social del conocimiento, el aprendizaje colaborativo en el aula virtual con sus particularidades, y las representaciones que se generan en los estudiantes a partir de las interacciones en el foro virtual.

Las actividades de aprendizaje se realizan de manera individual y/o grupal según se indique en las consignas. Con el fin de garantizar el debate y el intercambio se implementan foros con la coordinación del tutor/a, referidos a las temáticas de las unidades.

5.4 ANÁLISIS DE LA PRESENCIA COGNITIVA

5.4.1 Justificación de la unidad de análisis

Uno de los problemas que se presenta en el análisis del discurso asincrónico en entornos virtuales es la justificación de la unidad de análisis. Así, se considera el mensaje completo, el párrafo o la frase, desde el punto de vista textual; otras veces se opta por secuencias de contenido o temáticas, y otras, incluso, por estructuras gramaticales como pronombres, personas gramaticales, etc.

En los trabajos de algunos autores como Garrison, Anderson, Rourke, Archer y Gunawaderna se utiliza como unidad de análisis el mensaje completo, aunque conforme fueron desarrollando el modelo de análisis establecieron categorías e indicadores de menor aplicación discursiva, como la frase e incluso la palabra, pero siempre aplicados al mensaje completo como unidad principal de análisis.

Según Dewey, “la interacción es el componente que define el proceso educativo y tiene lugar cuando los estudiantes transforman la información inerte que se les transmite, en conocimiento con valor y aplicaciones personales” (Dewey, 1916 citado por Garrison y Anderson, 2005, p. 65). “La interacción entre profesores y estudiantes, la interacción entre los propios estudiantes y la colaboración en el aprendizaje que resulta de esas interacciones son elementos clave en un proceso educativo” (Pallof y Prat, 1999 citado por Garrison y Anderson, p. 66).

En este trabajo la unidad de análisis utilizada es el *mensaje completo* y se considera a éste cómo *el comprobante de la interacción*. Así como el comprobante de una compra-venta es la factura o el ticket en los cuales se detallan las cantidades y precios de los artículos involucrados en la transacción, en el mensaje se manifiestan los indicadores de la dimensión social, de la dimensión docente y de la dimensión cognitiva, involucrados en la interacción.

Cada mensaje se divide en partes que pueden ser identificadas en una de las subcategorías de análisis establecido. Un fragmento es una contribución interpretable y codificable en una de las subcategorías del sistema de análisis establecido (Coll Salvador, Bustos Sánchez y Engel Rocamora, 2011). El mensaje completo puede estar formado por uno o varios fragmentos.

La identificación de los fragmentos y su codificación se realiza por separado para cada una de las dimensiones social, docente y cognitiva. En lo que concierne al procedimiento de codificación, un mismo fragmento sólo puede ser codificado una vez en una subcategoría de cada dimensión. Los distintos fragmentos de un mensaje pueden ser codificados, si procede, en las distintas subcategorías de una misma dimensión, para más detalle ver Anexo B.

5.4.2. Heurísticas para la categorización

Para dotar de claridad al procedimiento de acuerdos y consensos en el desempeño de los codificadores e identificar las diferencias entre éstos, se realizan varias sesiones en la que se exponen las dudas y problemas del proceso de codificación. Fruto de éstas sesiones se diferencian y precisan cada una de las categorías de análisis de la dimensión cognitiva y se construye un instrumento con más detalle en la definición de cada código y un alto grado de exclusividad y exhaustividad entre los mismos.

El nuevo instrumento agiliza la tarea individual de los analistas y aumenta el nivel de acuerdo en las revisiones. En el Anexo B se muestra el instrumento construido, además se describe en forma detallada la heurística que permite obtener los datos preprocesados que son la base para la aplicación de la búsqueda de reglas de asociación y árboles de decisión.

Para mostrar la aplicación de la heurística e instrumentos desarrollados se selecciona uno de los grupos que participa del foro de aprendizaje del curso “Enseñar y Aprender en el aula virtual”. En la figura 5.3 se observa la presentación de la actividad grupal.

Para la preparación de los datos en primer lugar se compilan los mensajes del foro, se oculta la identidad de los participantes, tanto tutor como estudiantes, se asigna un número correlativo a cada mensaje y se genera un archivo.

El archivo generado se entrega a los docentes que van a codificar. La codificación se realiza de manera independiente por parte de tres analistas sobre los mismos textos utilizando la herramienta que se muestra en la tabla B.1.

Curso de Postgrado "Enseñar y Aprender en el Aula Virtual"

Foro Grupal Unidad I



Este es un espacio destinado a la construcción cognitiva grupal

Aquí estimularemos el desarrollo de capacidades y habilidades para analizar un tema de manera conjunta, la discusión, la capacidad para organizar los debates y escribir las conclusiones.

*Hemos organizado grupos de cuatro y cinco integrantes. **Acceda al enlace en el que esté incluido su apellido** y participe del debate sobre el tema planteado...*

Figura 5.3. Presentación de la actividad grupal de la Unidad I

La revisión de la codificación tiene lugar de manera compartida entre los investigadores que discuten sus dificultades para asignar cada fragmento a alguna categoría. En esta línea, durante las sesiones convocadas se utiliza un control de fiabilidad por consenso entre los tres investigadores, donde los registros se comparan para controlar el nivel de acuerdo existente. Se obtienen los datos preprocesados que son necesarios para aplicar las técnicas de minería de datos que se detallan en los siguientes incisos.

5.5. REGLAS DE ASOCIACIÓN

Como se expresa en el Capítulo 3, las reglas de asociación son una manera muy popular de expresar patrones de un conjunto de datos. En forma análoga a la aplicada a las cestas de compras, se trabaja con esta técnica para extraer conocimiento de los foros de aprendizaje.

Los diferentes mensajes de los foros son los comprobantes de la interacción que se genera en los mismos. Ellos se pueden expresar en una tabla, las filas de la tabla se refieren a los mensajes y las columnas son cada uno de indicadores de presencia cognitiva definidos. La tabla solo contiene valores binarios. Un 1 en la posición (i,j) indica que el mensaje i contiene el indicador j , mientras que un 0 indica que el mensaje no ha mostrado el indicador.

5.5.1. Definición

El problema de minería de reglas de asociación se define formalmente como:

Sea $I = \{i_1, i_2, \dots, i_n\}$ un conjunto de n atributos binarios llamados **ítems**.

Sea $D = \{t_1, t_2, \dots, t_m\}$ un conjunto de transacciones almacenadas en una base de datos.

Cada transacción en D tiene un **ID** (identificador) único y contiene un subconjunto de ítems de I .

Una **regla** se define como una implicación de la forma:

$$X \Rightarrow Y$$

Donde:

$$X, Y \subseteq I_y$$

$$X \cap Y = \emptyset$$

Los conjuntos de ítems X y Y se denominan respectivamente *antecedente* (o parte izquierda) y *consecuente* (o parte derecha) de la regla.

Para este caso: el conjunto I de los ítems son los indicadores de presencia cognitiva definidos en la tabla 1.

$I = \{\text{Sensación de perplejidad, Reconocer el problema, Confusión, \dots, Defender}\}$

Utilizando los identificadores queda:

$I = \{Ha, Hb, Hc, \dots, Rc\}$

El conjunto D de las transacciones es el conjunto de las interacciones y está representado por cada uno de los mensajes de los foros de los cursos objeto de este estudio.

$D = \{M_1, M_2, \dots, M_m\}$

El conjunto I tiene 15 atributos binarios llamados *indicadores de presencia cognitiva*, son binarios pues tiene dos valores posibles: estar o no en una transacción. Si el indicador está presente en un mensaje se le otorga el valor uno “1”; si no está se asignará el valor cero “0”.

El conjunto D contiene cada uno de los mensajes de los foros que se numeraron en forma correlativa, independientemente de su emisor, para lograr el ID (identificador) único de la

interacción. Este identificador cumple la función del número de factura o de ticket que se asigna en las transacciones de compra-venta.

Para el análisis de las reglas de asociación se separan las participaciones de los estudiantes, descartando las intervenciones de los tutores, y las que no contienen indicadores de presencia cognitiva ya que se trata sólo con los indicadores de esa dimensión, aún así se respeta el identificador único de cada mensaje. Es por eso que en las tablas se observa números de mensajes que no son correlativos.

A continuación se muestra la tabla de datos que corresponde a la codificación realizada para los mensajes del foro de la práctica de aprendizaje grupal de la unidad I, del curso de postgrado “Enseñar y Aprender en el Aula Virtual”. Esta práctica tiene 6 consignas y es presentada en la plataforma el domingo 7 de setiembre de 2008. Los resultados corresponden al grupo “G1”.

Consigna 1: Accedan a los Power Points y a los videos que acompañan el texto, a efectos de realizar un intercambio con sus compañeros de grupo, sobre las diferencias que presenta el abordaje de su propuesta de trabajo futuro, utilizando WEB 1.0 y WEB 2.0.

Tabla 5.1. *Indicadores de Presencia Cognitiva para los mensajes de la primera consigna*

	Ha	Hb	Hc	Hd	Ea	Eb	Ec	Ed	Ee	Ia	Ib	Ic	Ra	Rb	Rc
M1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
M3	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
M5	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0
M6	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0
M7	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
M10	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0

Consigna 2: En términos de la política educativa provincial e institucional ¿en qué etapa de incorporación de las TICs se encuentra la institución en la que cada uno de ustedes trabaja y/o conoce, según G. Sunkel?

Tabla 5.2 *Indicadores de Presencia Cognitiva para los mensajes de la segunda consigna*

	Ha	Hb	Hc	Hd	Ea	Eb	Ec	Ed	Ee	Ia	Ib	Ic	Ra	Rb	Rc
M3	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0
M5	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
M6	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
M9	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0

Consigna 3: Relate una experiencia en particular en el Foro de su grupo, e intercambie con sus compañeros acerca de las características más importantes que presenta cada una de ellas.

Tabla 5.3 Indicadores de Presencia Cognitiva para los mensajes de la tercera consigna

	Ha	Hb	Hc	Hd	Ea	Eb	Ec	Ed	Ee	Ia	Ib	Ic	Ra	Rb	Rc
M3	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
M5	0	1	0	0	0	1	0	0	0	0	1	1	0	0	0
M6	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
M8	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0

Consigna 4: Realicen un cuadro comparativo con las diferentes experiencias informáticas de los miembros del grupo y colóquenlo en este foro.

Tabla 5.4 Indicadores de Presencia Cognitiva para los mensajes de la cuarta consigna

	Ha	Hb	Hc	Hd	Ea	Eb	Ec	Ed	Ee	Ia	Ib	Ic	Ra	Rb	Rc
M3	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
M5	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
M6	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
M8	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
M11	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0

Consigna 5: ¿qué recursos tecnológicos para estudiantes, docentes/autoridades, dispone el establecimiento educativo en el que Ud. se desempeña? Y ¿a qué *densidad informática* atiende? Revisar el concepto de densidad informática en los textos seleccionados.

Tabla 5.5. Indicadores de Presencia Cognitiva para los mensajes de la quinta consigna

	Ha	Hb	Hc	Hd	Ea	Eb	Ec	Ed	Ee	Ia	Ib	Ic	Ra	Rb	Rc
M3	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
M5	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
M6	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
M8	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0

Consigna 6: Realicen grupalmente una síntesis de no más de dos páginas con las principales conclusiones en el foro del grupo.

Tabla 5.6. Indicadores de Presencia Cognitiva para los mensajes de la sexta consigna

	Ha	Hb	Hc	Hd	Ea	Eb	Ec	Ed	Ee	Ia	Ib	Ic	Ra	Rb	Rc
M11	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0

Siguiendo el procedimiento de codificación descrito en el Anexo B que dice:

A efectos del análisis de reglas de asociación, cuando en un mensaje hay varios fragmentos codificados en la misma categoría de una dimensión, se contabilizan una sola vez, se resumen los codificadores del grupo G1 y la codificación que resulta se muestra en la tabla 5.7.

Tabla 5.7. *Indicadores de Presencia Cognitiva Final Grupo 1*

	Ha	Hb	Hc	Hd	Ea	Eb	Ec	Ed	Ee	Ia	Ib	Ic	Ra	Rb	Rc
M1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
M3	1	1	0	1	0	1	0	1	0	0	0	0	0	0	0
M5	0	1	0	0	0	1	0	0	0	1	1	1	0	0	0
M6	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0
M7	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
M8	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
M9	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
M10	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
M11	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0

Se aplica la heurística para los otros grupos “G3” y “G5” y en resumen se tienen los siguientes datos:

Tabla 5.8 *Indicadores de Presencia Cognitiva Final 3 Grupos*

	GRUPO	IDENTIF.	Ha	Hb	Hc	Hd	Ea	Eb	Ec	Ed	Ee	Ia	Ib	Ic	Ra	Rb	Rc
1	G1-08	M1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
2	G1-08	M3	1	1	0	1	0	1	0	1	0	0	0	0	0	0	0
3	G1-08	M5	0	1	0	0	0	1	0	0	0	1	1	1	0	0	0
4	G1-08	M6	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0
5	G1-08	M7	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
6	G1-08	M8	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
7	G1-08	M9	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
8	G1-08	M10	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
9	G1-08	M11	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0
10	G3-11	M1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
11	G3-11	M2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
12	G3-11	M3	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
13	G3-11	M4	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
14	G3-11	M5	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
15	G3-11	M6	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
16	G3-11	M7	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
17	G3-11	M8	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0

18	G3-11	M9	0	1	0	0	0	1	0	0	0	0	0	0	0	0	
19	G3-11	M10	0	1	0	0	0	1	0	0	0	0	1	0	0	0	
20	G3-11	M11	0	0	0	0	0	1	0	1	0	0	1	0	0	0	
21	G3-11	M12	0	1	0	0	0	0	0	1	0	0	0	0	0	0	
22	G3-11	M13	0	0	0	0	0	0	0	0	0	1	0	0	0	0	
23	G3-11	M14	0	0	0	0	0	0	0	0	0	1	0	0	0	0	
24	G3-11	M15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
25	G3-11	M16	0	0	1	0	0	0	0	0	0	0	0	0	0	0	
26	G3-11	M17	0	0	0	0	0	1	0	0	0	1	1	0	0	0	
27	G5-11	M1	0	1	0	0	0	1	0	0	0	0	1	0	0	0	
28	G5-11	M3	0	1	0	0	0	1	1	0	0	0	1	0	0	0	
29	G5-11	M4	0	1	0	0	0	1	0	0	0	0	0	0	0	0	
30	G5-11	M5	0	0	0	0	0	1	0	0	0	0	0	0	0	0	
31	G5-11	M8	0	0	0	0	0	1	0	0	0	0	1	0	0	0	
32	G5-11	M10	0	0	0	0	0	0	0	0	0	1	0	0	0	0	
33	G5-11	M15	0	1	0	0	0	1	0	0	0	0	0	0	0	0	
34	G5-11	M16	0	0	0	0	0	0	0	0	0	1	0	0	0	0	
35	G5-11	M19	0	0	0	0	0	0	0	0	0	1	1	0	0	0	
36	G5-11	M20	0	0	0	0	0	0	0	0	0	1	0	0	0	0	
37	G5-11	M23	0	0	0	0	0	0	0	0	0	1	1	0	0	0	
38	G5-11	M24	0	0	0	0	0	0	0	0	0	1	0	0	0	0	
39	G5-11	M25	0	0	0	0	0	0	0	0	0	1	0	0	0	0	
40	G5-11	M28	0	0	0	0	0	0	0	0	0	1	0	0	0	0	
41	G5-11	M29	0	0	0	0	0	0	0	0	0	1	0	0	0	0	
42	G5-11	M30	0	0	0	0	0	0	0	0	0	1	0	0	0	0	
43	G5-11	M33	0	0	0	0	0	0	0	0	0	0	1	0	0	0	
44	G5-11	M34	0	0	0	0	0	0	0	0	0	1	0	0	0	0	
		TOTALES	1	19	3	2	0	19	1	3	0	17	12	2	0	0	
			Ha	Hb	Hc	Hd	Ea	Eb	Ec	Ed	Ee	Ia	Ib	Ic	Ra	Rb	Rc

5.5.2. Reglas significativas, 'soporte' y 'confianza'

Una *regla* se define como una implicación de la forma: $X \Rightarrow Y$

Donde:

$$X, Y \subseteq I_y$$

$$X \cap Y = \emptyset$$

Los conjuntos de ítems X y Y se denominan respectivamente *antecedente* (o parte izquierda) y *consecuente* (o parte derecha) de la regla.

En este caso por ejemplo la regla **Hb** \Rightarrow **Eb**

Significa que si el mensaje contiene Hb también contiene Eb, es decir, según la especificación formal anterior se tendría que:

X= {Todos los mensajes que contienen el indicador *Reconocer el problema*}

Y= {Todas los mensajes que contienen el indicador *Intercambio de información*}

O utilizando los identificadores:

X= {Todos los mensajes que contienen el indicador Hb}

Y= {Todas los mensajes que contienen el indicador Eb}

Tabla 5.9- *Mensajes con Hb*

GRUPO	IDENTIF.	Hb	Eb
G1-08	M3	1	1
G1-08	M5	1	1
G1-08	M6	1	1
G1-08	M8	1	1
G1-08	M9	1	1
G3-11	M1	1	0
G3-11	M2	1	0
G3-11	M3	1	0
G3-11	M4	1	1
G3-11	M6	1	0
G3-11	M7	1	1
G3-11	M8	1	1
G3-11	M9	1	1
G3-11	M10	1	1
G3-11	M12	1	0
G5-11	M1	1	1
G5-11	M3	1	1
G5-11	M4	1	1
G5-11	M5	0	1
G5-11	M8	0	1
G5-11	M15	1	1
	TOTALES	19	16

Para seleccionar reglas interesantes del conjunto de todas las reglas posibles se evalúan el *soporte* y la *confianza*. Estas dos medidas permiten conocer la calidad de la regla.

El soporte de un conjunto de ítems X en una base de datos D se define como la proporción de transacciones en la base de datos que contiene dicho conjunto de ítems:

$$\text{sop}(X) = \frac{|X|}{|D|}$$

En el ejemplo anterior el conjunto $D = \{\text{todos los mensajes de todas las consignas}\}$

El total de elementos del conjunto D o cardinal del conjunto es: $|D|= 44$.

$X= \{\text{Todos los mensajes que contienen el indicador Hb}\}$

El total de elementos del conjunto X o cardinal del conjunto es: $|X|= 19$

$$\text{sop}(X) = \frac{19}{44} = 0,43$$

Es decir, el soporte es del 43%. Esto significa que en el 43% de las interacciones de este foro se ha logrado “Reconocer el problema”.

La confianza de una regla se define como:

$$\text{conf}(X \Rightarrow Y) = \frac{\text{sop}(X \cup Y)}{\text{sop}(X)} = \frac{|X \cup Y|}{|X|}$$

Por ejemplo, para la regla: $\{\mathbf{Hb}\} \Rightarrow \{\mathbf{Eb}\}$

Tabla 5.10. Mensajes del conjunto $(\{Hb\} \cup \{Eb\})$

GRUPO	IDENTIF.	Hb	Eb
G1-08	M3	1	1
G1-08	M5	1	1
G1-08	M6	1	1
G1-08	M8	1	1
G1-08	M9	1	1
G3-11	M4	1	1
G3-11	M7	1	1
G3-11	M8	1	1
G3-11	M9	1	1
G3-11	M10	1	1
G5-11	M1	1	1
G5-11	M3	1	1
G5-11	M4	1	1

G5-11	M15	1	1
	TOTALES	14	14
		Hb	Eb

La confianza sería:

$$\text{Conf}(\{\mathbf{Hb}\} \Rightarrow \{\mathbf{Eb}\}) = \frac{\text{sop}(\{\mathbf{Hb}\} \cup \{\mathbf{Eb}\})}{\text{Sop}(\{\mathbf{Hb}\})} = \frac{|\mathbf{X} \cup \mathbf{Y}|}{|\mathbf{X}|} = \frac{14}{19} = 0,74$$

Este resultado significa que el 74% de las reglas de la base de datos que contienen 'Hb: Reconocer el problema' en el antecedente, tienen 'Eb: Intercambio de información' en el consecuente; en otras palabras, que la regla: $\{\mathbf{Hb}\} \Rightarrow \{\mathbf{Eb}\}$ es cierta en el 74% de los casos.

Se podría buscar otra regla, por ejemplo $\{\mathbf{Eb}\} \Rightarrow \{\mathbf{Ia}\}$

Tabla 5.11. Mensajes con Eb

GRUPO	IDENTIF.	Eb	Ia
G1-08	M3	1	0
G1-08	M5	1	1
G1-08	M6	1	0
G1-08	M8	1	0
G1-08	M9	1	0
G3-11	M4	1	0
G3-11	M5	1	0
G3-11	M7	1	0
G3-11	M8	1	1
G3-11	M9	1	0
G3-11	M10	1	0
G3-11	M11	1	0
G3-11	M17	1	1
G5-11	M1	1	0
G5-11	M3	1	0
G5-11	M4	1	0
G5-11	M5	1	0
G5-11	M8	1	0
G5-11	M15	1	0
	TOTALES	19	17
		Eb	Ia

$$\text{sop}(X) = \frac{19}{44} = 0,43$$

El mismo soporte que la primera regla.

Tabla 5.12. Mensajes del conjunto ($\{Eb\} \cup \{Ia\}$)

GRUPO	IDENTIF.	Eb	Ia
G1-08	M5	1	1
G3-11	M8	1	1
G3-11	M17	1	1
	TOTALES	3	3

$$\text{Conf}(\{Eb\} \Rightarrow \{Ia\}) = \frac{3}{19} = 0,16$$

La confianza es mucho más baja.

Las reglas de asociación deben satisfacer las especificaciones del usuario en cuanto a umbrales mínimos de soporte y confianza. Para conseguir esto el proceso de generación de reglas de asociación se realiza en dos pasos. Primero se aplica el soporte mínimo para encontrar los conjuntos de ítems más frecuentes en la base de datos. En segundo lugar se forman las reglas partiendo de estos conjuntos frecuentes de ítems y de la restricción de confianza mínima. Encontrar todos los subconjuntos frecuentes de la base de datos es difícil ya que esto implica considerar todos los posibles subconjuntos de ítems o combinaciones de ítems. El conjunto de posibles conjuntos de ítems es el conjunto potencia de I y su tamaño es de $2^n - 1$). En este caso a pesar de que el conjunto de ítems tiene sólo 15 elementos, es decir habría que considerar $2^{15} - 1 = 32767$ conjuntos, excluyendo el conjunto vacío que no es válido como conjunto de ítems. Por este motivo se utilizan productos software que automaticen los procesos. En el Anexo A se describe uno de ellos.

5.5.3. Aplicaciones

En uno de los trabajos presentados en el Capítulo 4, la autora Ida Fallas Monge (2008) expresa “los estudiantes que alcanzan la fase de resolución, también tienen mensajes en la fase de integración, y además no tienen ningún mensaje en la fase inicial correspondiente a la categoría de hecho desencadenante”. Es decir, posiblemente hay una regla de asociación entre los indicadores de presencia cognitiva, pero no se muestran los datos que permitan confirmarlo.

La mayoría de los trabajos investigados calculan los porcentajes de mensajes que están en cada una de las categorías y generan un ranking que ordena las categorías según esos porcentajes. Pero no analizan las posibles relaciones entre las categorías, si hay alguna que promueva alcanzar otras, o cuáles afectan negativamente.

Por ejemplo, la regla *Reconocer el problema* => *Intercambio de información* con un soporte del 74%, está indicando que el *No reconocer el problema* también afectará al *Intercambio de información*.

El tutor puede utilizar los indicadores observados para reorientar el desarrollo de la discusión; ya que las reglas ofrecen a los docentes un medio para *evaluar el proceso de aprendizaje* y detectar si se necesitan realizar ajustes.

El descubrimiento de reglas de asociación se aplica con éxito en sistemas de comercio electrónico, para comprender el comportamiento de los clientes y poder incrementar las ventas. En e-learning se trata de extraer conocimiento de las relaciones entre los indicadores de presencia cognitiva que permitan conocer el comportamiento de los estudiantes con el objetivo de guiarlos durante su aprendizaje para maximizarlo

5.6. ÁRBOLES DE DECISIÓN

5.6.1. Nivel de Presencia Cognitiva

Un árbol de decisión es un conjunto de condiciones organizadas en una estructura jerárquica, de tal manera que la decisión final a tomar se puede determinar siguiendo las condiciones que se cumplen desde la raíz del árbol hasta alguna de sus hojas. En esta parte del trabajo se aplicarán distintas técnicas de clasificación, que se analizaron en el Capítulo 3, con el objetivo de construir árboles de decisión que describan el nivel de presencia cognitiva alcanzado por los estudiantes en los foros de aprendizaje de los cursos.

Los algoritmos que se utilizan están enmarcados en el proceso de reglas de inducción automatizadas. Inducción significa un proceso de razonamiento desde lo específico a lo general. A un sistema de inducción se le dan ejemplos del problema a tratar, llamado conjunto de entrena-

miento donde la salida se conoce. Después que se suministran varios ejemplos, el sistema de inducción genera reglas que sintetizan los casos cargados.

A partir de los indicadores de presencia cognitiva observados en los foros de aprendizaje, se pretende deducir el conocimiento existente y finalmente expresarlo en la menor cantidad de reglas posibles.

En forma análoga a la aplicada para las reglas de asociación, los valores de los indicadores de presencia cognitiva observados en los foros de aprendizaje se organizan en una tabla. En este caso, cada fila representa *un estudiante* que participa en los foros de aprendizaje, las columnas son los indicadores de presencia cognitiva a alcanzar *para un objetivo de aprendizaje*. La tabla sólo contiene valores binarios. Un 1 en la posición (i,j) indica que el estudiante i ha alcanzado el indicador j, mientras que un 0 indica que el estudiante no ha logrado el indicador.

Para completar la tabla resta definir un resultado o *función objetivo* que se obtiene del conjunto de valores observados, de acuerdo con un dominio de conocimiento que se quiere adquirir. Este resultado para el dominio que se está analizando es el *Nivel de Presencia Cognitiva* (NPC) alcanzado por cada estudiante para un objetivo de aprendizaje, y tendrá los valores: *Muy Alto, Alto, Medio o Bajo*.

En un ambiente de enseñanza en línea, el docente debe planificar las actividades de aprendizaje y los objetivos que deberán alcanzar los estudiantes con cada una de ellas. Dichas actividades permitirán que el estudiante construya significados mediante la reflexión continua en una comunidad de aprendizaje, es decir logrará alcanzar distintos niveles de presencia cognitiva.

En base a los objetivos, las actividades y los criterios de evaluación contenidos en la planificación de la práctica de aprendizaje grupal de la unidad I, el equipo docente construye el conjunto de entrenamiento para la clasificación del nivel de presencia cognitiva, que se detalla en la tabla 5.13.

Tabla 5.13 *Conjunto de entrenamiento para la clasificación*

Ha	Hb	Hc	Hd	Ea	Eb	Ec	Ed	Ee	Ia	Ib	Ic	Ra	Rb	Rc	NPC
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	bajo
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	bajo

0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	bajo
0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	bajo
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	bajo
0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	bajo
0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	bajo
0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	bajo
0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	bajo
0	1	0	0	0	0	0	1	0	1	0	0	0	0	0	medio
0	1	0	0	0	1	0	0	0	1	0	0	0	0	0	medio
0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	medio
0	1	0	0	0	1	1	0	0	1	0	0	0	0	0	medio
0	1	0	0	0	1	0	1	0	1	0	0	0	0	0	medio
0	1	0	0	0	1	0	0	0	1	1	0	0	0	0	alto
0	1	0	0	0	1	1	0	0	0	1	1	0	0	0	alto
0	1	0	0	0	1	0	0	0	1	1	0	0	0	0	alto
1	1	0	0	0	1	0	0	0	0	1	1	0	0	0	alto
1	1	0	0	0	1	0	0	0	1	1	1	0	0	0	alto
0	1	0	0	0	1	0	0	0	0	1	1	1	0	0	muyalto
1	1	0	0	0	1	0	0	0	0	1	1	1	1	0	muyalto
1	1	0	0	0	1	0	0	0	0	1	1	1	1	1	muyalto
1	1	0	0	0	1	1	1	1	1	1	1	1	1	1	muyalto

Para construir el conjunto a clasificar se utiliza la misma codificación de los mensajes que se usa para el análisis de las reglas de asociación, pero reuniendo todas las participaciones de un mismo estudiante que correspondan al mismo objetivo de aprendizaje. Es decir, si el estudiante E1 es el emisor de los mensajes M1, M10, M16 se trasladan a la nueva tabla los valores de los indicadores obtenidos en esos mensajes. La tabla 5.14 muestra los datos que corresponde a la codificación realizada para los mensajes del foro de la práctica de aprendizaje grupal de la unidad I, del curso de postgrado *Prácticas educativas en Entornos Virtuales*. Estos indicadores corresponden a los cuatro integrantes del grupo G5. En el Anexo B se muestra el proceso completo. En la figura 5.4 se observa la presentación de la práctica y en la figura 5.5 las consignas de la práctica grupal.

A trabajar en grupo

Estimad@s estudiantes: A continuación debajo de "Temas", cada uno de los grupos encontrará su espacio destinado al trabajo colaborativo grupal, *identificado con sus nombres*.

En él se presenta la *consigna* correspondiente a la primera práctica de aprendizaje. Es necesario entonces que identifiquen a sus compañer@s y se comuniquen para comenzar a interactuar.

¡Bienvenidos al foro y buen trabajo!!

Tema	Comenzado por	Respuestas	Último mensaje
------	---------------	------------	----------------

Figura 5.4 Presentación práctica de aprendizaje grupal de la unidad I - 2011

Prácticas Educativas en Entornos Virtuales

A partir de la lectura de los materiales seleccionados para esta Unidad es necesario que:

- interactúen con su grupo acerca de **los procesos de enseñanza y aprendizaje en entornos virtuales** y complementenlo desde su experiencia como profesores,
- expliciten las particularidades de los entornos virtuales para diseñar propuestas didácticas en la modalidad presencial, semipresencial y/o no-presencial.

En relación a las conclusiones del debate:

- escribir las conclusiones como producto de todo el equipo,
- mencionar los acuerdos y las diferencias entre los miembros,
- adjuntar el resumen (tres páginas) en este foro en la fecha estipulada

Fecha tope de confección de síntesis de lo producido grupalmente en respuesta a la Práctica de Aprendizaje: **2 de mayo**.

[Editar](#) | [Borrar](#) | [Responder](#)

Figura 5.5 Consignas práctica de aprendizaje grupal de la unidad I

Tabla 5.14. *Indicadores de Presencia Cognitiva Grupo 5*

	Ha	Hb	Hc	Hd	Ea	Eb	Ec	Ed	Ee	Ia	Ib	Ic	Ra	Rb	Rc	NPC
E1	0	1	0	0	0	1	0	0	0	1	1	0	0	0	0	¿?
E2	0	1	0	0	0	1	1	0	0	1	1	0	0	0	0	¿?
E3	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0	¿?
E4	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	¿?

Se detallan los indicadores y valores de presencia cognitiva obtenidos por el estudiante E1 para la práctica de aprendizaje grupal de la unidad I, se muestran en la tabla 5.15.

Tabla 5.15 *Indicadores de Presencia Cognitiva Estudiante E1*

Categoría	Indicador	Valor	Identificador
HECHO DESENCADENANTE	Sensación de perplejidad	0	Ha
	Reconocer el problema	1	Hb
	Confusión	0	Hc

	Contingencia	0	Hd
EXPLORACIÓN	Divergencia	0	Ea
	Intercambio de información	1	Eb
	Sugerencias	0	Ec
	Lluvia de ideas	0	Ed
	Saltos intuitivos	0	Ee
INTEGRACIÓN	Convergencia	1	Ia
	Síntesis	1	Ib
	Soluciones	0	Ic
RESOLUCIÓN	Aplicar	0	Ra
	Comprobar	0	Rb
	Defender	0	Rc

Es decir, el estudiante E1 ha logrado para el objetivo de aprendizaje que se analiza:

- Localizar y delimitar el problema y exponer sus ideas fundamentadas en la bibliografía propuesta para la consigna.
- Aportar documentos, direcciones electrónicas, distinto tipo de información o compartir experiencias que agregan conocimiento sobre la temática en relación con la consigna.
- Analizar las ideas propuestas, manifestar acuerdos con otras contribuciones y organizar la información progresivamente.
- Combinar las ideas propuestas para formar un todo coherente tratando de construir una explicación al problema planteado.

Se trata entonces, de definir en base a los indicadores observados el *Nivel de Presencia Cognitiva* alcanzado para ese objetivo de aprendizaje. Para las prácticas grupales de los casos analizados, la definición del *Nivel de Presencia Cognitiva* se realiza de manera independiente por tres de los tutores sobre los mismos estudiantes. La revisión de los resultados asignados tiene lugar de manera compartida entre los tutores. Durante las sesiones de revisión se utiliza un control de fiabilidad por consenso entre los tres docentes, donde los registros son contrastados para controlar el nivel de acuerdo existente. Los indicadores de la tabla 5.16 corresponden a los integrantes de los G3 y G5.

Tabla 5.16. *Indicadores y Nivel de Presencia Cognitiva Grupos G3 y G5.*

	Ha	Hb	Hc	Hd	Ea	Eb	Ec	Ed	Ee	Ef	Ia	Ib	Ic	Id	Ra	Rb	Rc	Rd	NPC
E1	0	1	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	Alto
E2	0	1	0	0	0	1	1	0	0	0	1	1	0	0	0	0	0	0	Alto
E3	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	Medio

E4	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	Alto
E5	0	1	0	1	0	1	0	1	0	0	1	1	0	0	0	0	0	0	Alto
E6	0	1	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	Medio
E7	0	1	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0	Medio
E8	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	Bajo

5.6.2. Algoritmo J48

En este trabajo se utiliza el software WEKA como herramienta de prueba de los algoritmos de aprendizaje. Como se explica en el Capítulo 3, J48 es una implementación del algoritmo C4.5 en la herramienta WEKA para minería de datos. En el Anexo A se explica el funcionamiento básico de esta herramienta.

Weka trabaja principalmente sobre archivos con un formato denominado arff, acrónimo de *Attribute-Relation File Format*. Este formato está compuesto por una estructura claramente diferenciada en tres partes:

1. **Cabecera.** En esta sección se define el nombre de la relación. Su sintaxis es la siguiente:

@relation <nombre-de-la-relación>

Se define el nombre de la relación como “presenciacognitiva”.

2. **Declaraciones de atributos.** En esta sección se declaran los atributos que compondrán el archivo junto a su tipo, su sintaxis es la siguiente:

@attribute <nombre-del-atributo> <tipo>

Los atributos son los indicadores de presencia cognitiva. Los valores posibles son 0 y 1.

Los nombres de los atributos se definen con la siguiente sintaxis: las dos primeras letras del nombre indican la categoría y el resto es el nombre completo de la subcategoría.

Por ejemplo el atributo *hdsensaciondeperplejidad* las dos primeras letras del nombre *hd* indican la categoría *hecho desecandante* y *sensación de perplejidad* es el nombre de la subcategoría.

También se define como atributo la función objetivo *nivelpresenciacognitiva* y sus valores; son: muyalto, alto, medio, bajo.

3. **Sección de datos.** Se declaran los datos que componen cada transacción separando entre comas los atributos y con saltos de línea las relaciones. En este caso son los valores de los indicadores de presencia cognitiva.

@data

Los datos de la tabla 5.9 se transforman al formato que utiliza WEKA y queda un archivo con extensión *.arff* con la siguiente estructura:

@relation presenciacognitiva

@attribute hdsensaciondeperplejidad {0,1}

@attribute hdreconocerelproblema {0,1}

@attribute hdconfusion {0,1}

@attribute hdotros {0,1}

@attribute exdivergencia {0,1}

@attribute exintercambiodeinformacion {0,1}

@attribute exsugerencias {0,1}

@attribute exlluviadeideas {0,1}

@attribute exsaltosintuitivos {0,1}

@attribute inconvergencia {0,1}

@attribute insintesis {0,1}

@attribute insoluciones {0,1}

@attribute reaplicar {0,1}

@attribute recomprobar {0,1}

@attribute redefender {0,1}

@attribute nivelpresenciacognitiva {muyalto, alto, medio, bajo}

@data

0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,bajo

1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,bajo

0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,bajo

0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,bajo

0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,bajo

0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,bajo

0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,bajo

0,1,0,0,0,1,0,0,0,0,0,0,0,0,0,bajo

0,1,0,0,0,0,1,0,0,0,0,0,0,0,0,bajo

0,1,0,0,0,0,0,1,0,1,0,0,0,0,0,medio

0,1,0,0,0,1,0,0,0,1,0,0,0,0,0,medio

0,1,0,0,0,0,0,0,0,1,0,0,0,0,0,medio

0,1,0,0,0,1,1,0,0,1,0,0,0,0,0,medio

0,1,0,0,0,1,0,1,0,1,0,0,0,0,0,medio

0,1,0,0,0,1,0,0,0,1,1,0,0,0,0,alto
0,1,0,0,0,1,1,0,0,0,1,1,0,0,0,alto
0,1,0,0,0,1,0,0,0,1,1,0,0,0,0,alto
1,1,0,0,0,1,0,0,0,0,1,1,0,0,0,alto
1,1,0,0,0,1,0,0,0,1,1,1,0,0,0,alto
0,1,0,0,0,1,0,0,0,0,1,1,1,0,0,muyalto
1,1,0,0,0,1,0,0,0,0,1,1,1,1,0,muyalto
1,1,0,0,0,1,0,0,0,0,1,1,1,1,1,muyalto
1,1,0,0,0,1,1,1,1,1,1,1,1,1,1,muyalto

De las opciones provistas por Weka en el selector de interfases, se trabaja con el Explorador.

De las seis alternativas que permite el explorador se selecciona la tarea de clasificación con el algoritmo J48. (Figura 5.6)

Una vez elegido el clasificador el próximo paso es la configuración del modo de evaluación del clasificador. El resultado de aplicar el algoritmo de clasificación se efectúa comparando la clase predicha con la clase real de las instancias. Esta evaluación puede realizarse de diferentes modos, según la selección en el cuadro *Test options* que se muestra en la figura 5.7. Weka proporciona 4 modos de prueba: *Use training set*, *Supplied test set*, *Cross-validation*, *Percentage Split*.

5.6.2.1. Algoritmo J48 con modo Use training set

Para la primera clasificación se selecciona el modo de evaluación Use training set, y se ejecuta el clasificador figura 5.7.

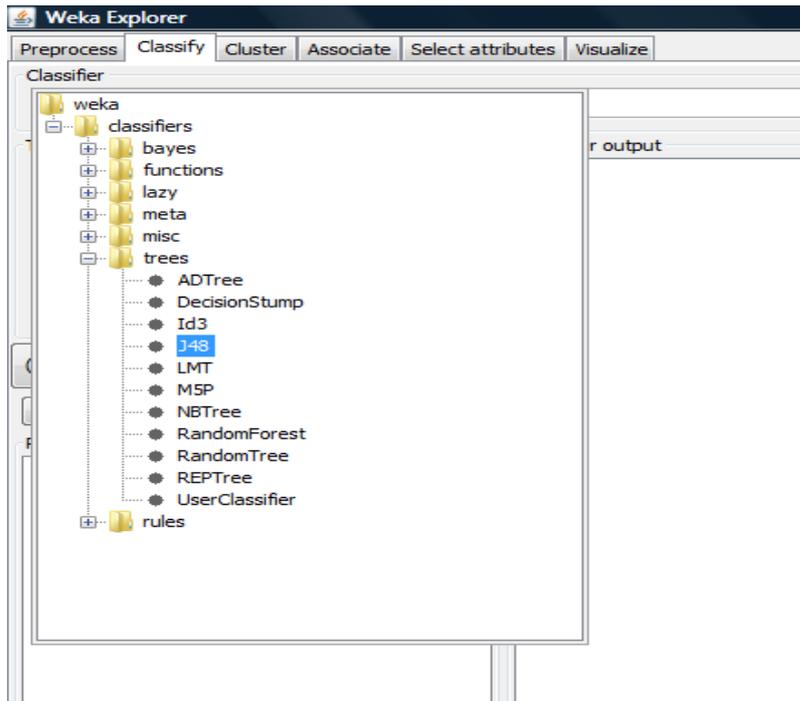


Figura 5.6 Selección del algoritmo J48

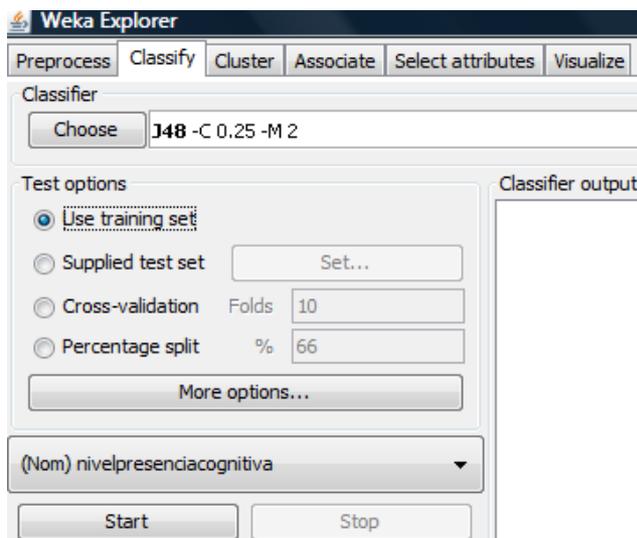


Figura 5.7. Selección del modo de evaluación o Test Options

Una vez que se ejecuta el clasificador seleccionado, en la ventana de texto de la derecha aparece la salida del clasificador o *classifiers output* que es un listado con los resultados, dónde se detalla: *run information*, *classifier model* y *evaluation*.

Información de ejecución o *run information* detalla el algoritmo utilizado *weka.classifiers.trees.J48 -C 0,25 -M 2*, el nombre de la relación, en este caso *presenciacognitiva*, la cantidad de

instancias o filas procesadas, 23, la cantidad de atributos 16, los nombres de cada atributo y el modo en que se ejecuta la clasificación o *evaluate on training data*, la imagen se muestra en la figura 5.8.

```

Classifier output
=== Run information ===

Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    presenciacognitiva
Instances:   23
Attributes:  16
             hdsensaciondeperplejidad
             hdreconocereproblema
             hdconfusion
             hdotros
             exdivergencia
             exintercambiodeinformacion
             eksugerencias
             exlluviadeideas
             exsaltosintuitivos
             inconvergencia
             insintesis
             insoluciones
             reaplicar
             recomprobar
             redefender
             nivelpresenciacognitiva
Test mode:   evaluate on training data

```

Figura 5.8. Información de ejecución.

Modelo de clasificación o *classifier model* indica el modo en que se ejecuta el clasificador “full training set” en este caso y detalla las reglas que permiten clasificar cada hoja del árbol. Es interesante observar que de los 16 atributos iniciales aparecen sólo 3 atributos en las reglas que genera. En el inciso 5.6.3 se explican estas reglas (figura 5.9).

```

=== Classifier model (full training set) ===

J48 pruned tree
-----

insintesis = 0
|  inconvergencia = 0: bajo (9.0)
|  inconvergencia = 1: medio (5.0)
insintesis = 1
|  reaplicar = 0: alto (5.0)
|  reaplicar = 1: muyalto (4.0)

Number of Leaves :    4
Size of the tree :    7

Time taken to build model: 0.05 seconds

```

Figura 5.9. Modelo de clasificación

5.6.2.2. Evaluación del clasificador

Con referencia al informe de evaluación del clasificador, se pueden destacar tres elementos: *Summary, Detailed Accuracy By Class, Confusion Matrix*.

```

=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances      23      100    %
Incorrectly Classified Instances    0        0    %
Kappa statistic                    1
Mean absolute error                 0
Root mean squared error             0
Relative absolute error              0    %
Root relative squared error         0    %
Total Number of Instances          23

```

Figura 5.10 *Resumen*

Resumen o *Summary*: es el porcentaje global de errores cometidos en la evaluación. En los datos de medición de error hay tres campos: instancias correctamente clasificadas, instancias incorrectamente clasificadas y número total de instancias. Son las instancias modeladas correctamente, incorrectamente y el total modelado. Se puede observar en las figuras 5.10 y 5.17 que los resultados asociados a medición de errores marcan que no existe error en la clasificación de las instancias.

El estadístico Kappa, coeficiente Kappa o *Kappa Statistic* se usa como un significado del grado de clasificación en datos categóricos. Un valor en el coeficiente Kappa de 1 significa un modelado estadísticamente perfecto, mientras un valor de 0 significa que cada valor del modelo fue diferente al valor actual o real. Un coeficiente kappa 0.7 o mayor es lo anhelado como correlación estadística Buena, pero por supuesto, el valor más alto da la mejor correlación.

Error absoluto medio o *Root mean-squared error*: es el promedio de la diferencia entre el valor predicho y el actual de testeo para todos los casos de testeo. Es el error de predicción promedio.

Error cuadrático relativo o *Root relative squared error*: es el total de los errores cuadráticos relativo a los errores cuadráticos respecto del valor medio absoluto. En este caso se exagera el error en aquellos casos en que el error de predicción es significativamente mayor que el error medio.

Error absoluto relativo o *Relative absolute error* es el error absoluto total relativo al error que debería existir si la predicción se hubiera realizado como el promedio de los valores actuales.

```

==== Detailed Accuracy By Class ====
TP Rate  FP Rate  Precision  Recall  F-Measure  Class
1         0         1           1         1           muyalto
1         0         1           1         1           alto
1         0         1           1         1           medio
1         0         1           1         1           bajo

```

Figura 5.11. *Precisión detallada por clase*

Precisión detallada por clase o *Detailed Accuracy By Class*: para cada uno de los valores que puede tomar el atributo de clase: el porcentaje de instancias con ese valor que son correctamente predichas (TP: true positives), y el porcentaje de instancias con otros valores que son incorrectamente predichas a ese valor aunque tenían otro (FP: false positives). Las otras columnas, *precision*, *recall*, *F-measure*, se relacionan con estas dos anteriores. (figura 5.11)

Matriz de confusión o *Confusion Matrix*: Es una herramienta de visualización que se emplea en aprendizaje supervisado. Cada columna de la matriz representa el número de predicciones de cada clase, mientras que cada fila representa a las instancias en la clase real. Uno de los beneficios de las matrices de confusión es que facilitan ver si el sistema está confundiendo dos clases. Los valores de la diagonal son los aciertos, y el resto de los valores son los errores, se puede observar en las figuras 5.12 y 5.19.

```

==== Confusion Matrix ====
a b c d  <-- classified as
4 0 0 0 | a = muyalto
0 5 0 0 | b = alto
0 0 5 0 | c = medio
0 0 0 9 | d = bajo

```

Figura 5.12. *Matriz de confusión*

Al igual que con otras opciones de análisis, la ventana izquierda de la lista de resultados contiene el resumen de todas las aplicaciones de clasificadores sobre conjuntos de datos en la sesión del *Explorer*. Puede accederse a esta lista para presentar los resultados, y al activar el botón derecho aparecen diferentes opciones de visualización, que se muestran en la figura 5.13.

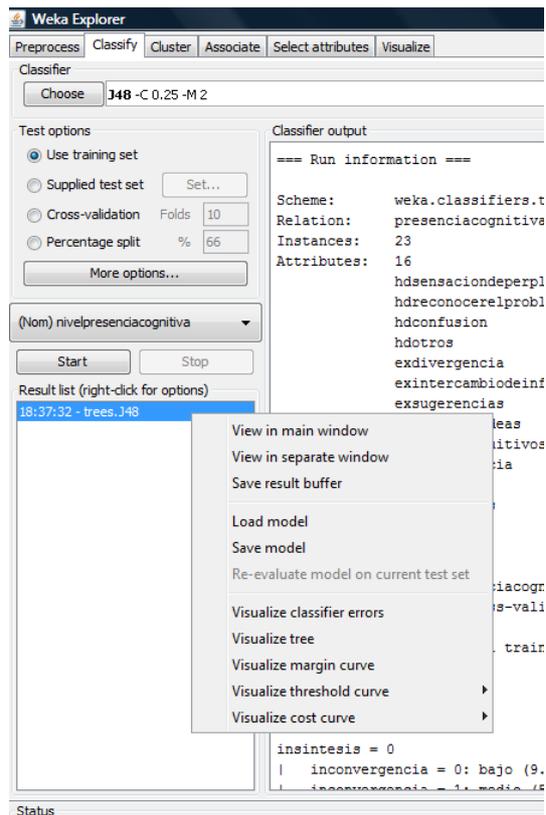


Figura 5.13. Lista de resultados

Si se selecciona la opción *Visualize tree*, se puede ver el árbol que generó el clasificador (figura 5.14). En cada nodo del árbol, el algoritmo J48 elige el atributo de los datos que más eficazmente divide el conjunto de instancias en subconjuntos. Su criterio es el normalizado para ganancia de información (diferencia de entropía) que resulta en la elección de un atributo para dividir los datos. El atributo con la mayor ganancia de información normalizada se elige como parámetro de decisión (ver Capítulo 3).

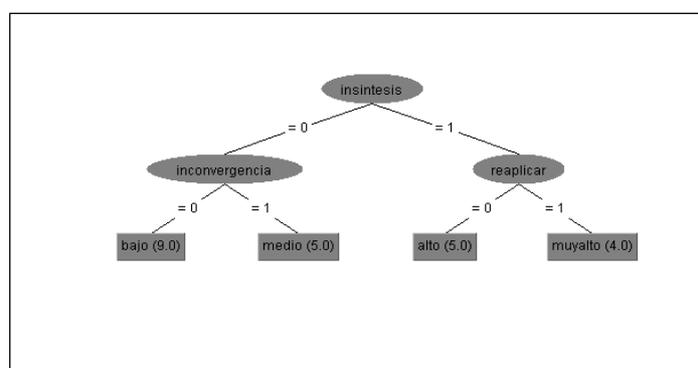


Figura 5.14. Árbol de clasificación

5.6.2.3. Algoritmo J48 con modo Supplied test

Se realiza otra clasificación seleccionando el modo Supplied test. Para ejecutarla se utiliza la tabla 5.12 correspondiente a los indicadores de presencia cognitiva observados en dos grupos de estudiantes, que en formato .arff es el siguiente:

```
@relation presenciacognitiva
@attribute hdsensaciondeperplejidad {0,1}
@attribute hdreconocerelproblema {0,1}
@attribute hdconfusion {0,1}
@attribute hdotros {0,1}
@attribute exdivergencia {0,1}
@attribute exintercambiodeinformacion {0,1}
@attribute exsugerencias {0,1}
@attribute exlluviadeideas {0,1}
@attribute exsaltosintuitivos {0,1}
@attribute inconvergencia {0,1}
@attribute insintesis {0,1}
@attribute insoluciones {0,1}
@attribute reaplicar {0,1}
@attribute recomprobar {0,1}
@attribute redefender {0,1}
@attribute nivelpresenciacognitiva {muyalto,alto, medio, bajo}
@data
0,0,0,1,0,0,0,1,0,1,1,0,0,0,0,alto
0,0,0,1,0,0,0,0,0,1,0,0,0,0,0,medio
0,1,0,1,0,1,0,1,0,1,0,0,0,0,0,medio
0,1,0,1,0,1,0,0,0,0,0,0,0,0,0,bajo
0,0,0,0,0,1,0,1,0,1,0,0,0,0,0,medio
0,0,0,0,0,1,0,0,0,1,0,0,0,0,0,medio
0,0,0,0,0,1,1,1,0,1,0,0,0,0,0,medio
0,0,0,0,0,1,1,1,0,1,1,0,0,0,0,alto
```

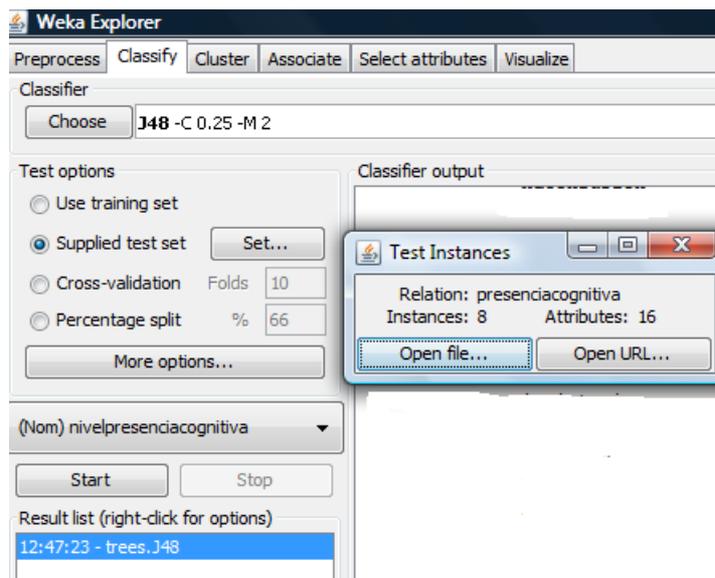


Figura 5.15. Selección del modo *Supplied test set*

Una vez que se ejecuta el clasificador con el nuevo modo seleccionado, aparece en la ventana de texto de la derecha la salida del clasificador que se muestra en las figuras 5.16 a 5.20.

```
=== Run information ===

Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    presenciacognitiva
Instances:   23
Attributes:  16
             hdsensaciondeperplejidad
             hdreconocerelproblema
             hdconfusion
             hdotros
             exdivergencia
             exintercambiodeinformacion
             eksugerencias
             exlluviadeideas
             exsaltosintuitivos
             inconvergenca
             insintesis
             insoluciones
             reaplicar
             recomprobar
             redefender
             nivelpresenciacognitiva
Test mode:   user supplied test set: 8 instances
```

Figura 5.16. Información de ejecución.

```
=== Classifier model (full training set) ===

J48 pruned tree
-----

insintesis = 0
|  inconvergenca = 0: bajo (9.0)
|  inconvergenca = 1: medio (5.0)
insintesis = 1
|  reaplicar = 0: alto (5.0)
|  reaplicar = 1: muyalto (4.0)

Number of Leaves :    4

Size of the tree :    7

Time taken to build model: 0 seconds
```

Figura 5.17. Modelo de clasificación

```

=== Evaluation on test set ===
=== Summary ===

Correctly Classified Instances      8          100    %
Incorrectly Classified Instances    0           0    %
Kappa statistic                     1
Mean absolute error                 0
Root mean squared error             0
Relative absolute error              0    %
Root relative squared error         0    %
Total Number of Instances          8

```

Figura 5.18 *Resumen*

```

=== Detailed Accuracy By Class ===
TP Rate  FP Rate  Precision  Recall  F-Measure  Class
0        0        0          0        0          muyalto
1        0        1          1        1          alto
1        0        1          1        1          medio
1        0        1          1        1          bajo

```

Figura 5.19. *Precisión detallada por clase*

```

=== Confusion Matrix ===
 a b c d  <-- classified as
0 0 0 0 | a = muyalto
0 2 0 0 | b = alto
0 0 5 0 | c = medio
0 0 0 1 | d = bajo

```

Figura 5.20. *Matriz de confusión*

La clasificación del algoritmo J48 con dos de sus modos, Use training set, Supplied test set muestra resultados óptimos con el conjunto de datos utilizados. Se comprueba con datos de las categorizaciones de los análisis individuales de los codificadores y detecta los mismos errores que fueron corregidos durante la etapa de revisión, es decir el clasificador se adapta al dominio de conocimiento que se quiere adquirir.

5.6.3. Reglas

Del árbol obtenido con el algoritmo J48 se obtienen cuatro reglas, que pueden ser utilizadas por los docentes para la evaluación del nivel de presencia cognitiva de cada una de las actividades de los foros, sin perder información. O pueden ser almacenadas en una computadora por medio de cualquier lenguaje:

REGLA1

SI (insintesis = 0) AND (inconvergencia = 0) ENTONCES (nivelpresenciacognitiva= bajo)

REGLA2

SI (insíntesis = 0) AND (inconvergenca = 1) **ENTONCES** (nivepresenciacognitiva= medio)

REGLA3

SI (insíntesis = 1) AND (reaplicar = 0) **ENTONCES** (nivepresenciacognitiva= alto)

REGLA4

SI (insíntesis = 1) AND (reaplicar = 1) **ENTONCES** (nivepresenciacognitiva= muyalto)

La categoría *Integración*, con sus indicadores *convergenca* y *síntesis* permite evaluar el nivel de presencia cognitivo **bajo** y **medio**.

La categoría *Integración* con su indicador *síntesis* en conjunción con la categoría *Resolución* con su indicador *aplicar* permite evaluar el nivel de presencia cognitivo **alto** y **muy alto**.

Es decir estas dos categorías son las que brindan mayor información respecto al nivel de presencia cognitiva.

5.7. MEJORA CONTINUA

En el Capítulo 2 se describe el ciclo de Deming o ciclo de calidad que propone una forma de acción constante en pos del mejoramiento continuo. Plantea cuatro etapas: planear, hacer, verificar y actuar. *Planear* incluye determinar metas, objetivos y métodos. *Hacer* implica dar educación y capacitación a los involucrados y realizar el trabajo en sí. *Verificar* consiste en evaluar los resultados de lo hecho. *Actuar* implica tomar medidas correctivas.

Las reglas que brindan mayor información respecto al nivel de presencia cognitiva permiten realizar la evaluación de los resultados de aprendizaje y a partir de allí establecer medidas correctivas. Es decir contribuye a *verificar* y *actuar*. Pero además, sirven de fundamento para anticiparse y tomar acciones de carácter preventivo, es decir *planear*.

Para que los estudiantes alcancen un nivel de presencia cognitivo alto o muy alto, una forma sería incluir en las propuestas de aprendizaje de los foros actividades que enfatizen las fases

de integración y resolución, es decir, que promuevan el logro de los indicadores de las categorías que brindan mayor información para esa dimensión.

CAPÍTULO 6

Conclusiones y futuras líneas de investigación

CAPÍTULO 6: CONCLUSIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN

6.1. SÍNTESIS: EL PROBLEMA Y LA SOLUCIÓN

Evaluar la calidad o éxito en el proceso de aprendizaje en un entorno e-learning es una labor muy compleja. El modelo de investigación práctica propuesto por Garrison y Anderson (2005) permite evaluar la presencia cognitiva, además ofrece un medio para juzgar la naturaleza y calidad de la reflexión y el discurso crítico en una comunidad de investigación que trabaja en cooperación.

La presencia cognitiva es una condición del pensamiento y el aprendizaje de alto nivel y se refiere a los resultados educativos pretendidos y conseguidos (Garrison y Anderson, 2005). Es decir, la presencia cognitiva es *“el punto hasta el cual los estudiantes son capaces de construir significado mediante la reflexión continua en una comunidad de investigación”* (Garrison, Anderson y Archer 2001 citado en Garrison y Anderson, 2005, p. 50).

Estos autores presentan un conjunto de descriptores e indicadores de la presencia cognitiva que son aptos para valorar la calidad de la investigación. Se trata de emplear esos indicadores para evaluar el nivel de pensamiento y discurso críticos en el gran número de mensajes que se generan durante las sesiones de e-learning.

Debido a la importancia central de la interacción en un entorno de educación no presencial, se considera relevante analizar el papel que desempeñan los foros en los entornos virtuales, ya que éstos proveen el principal espacio para el intercambio académico.

Existen numerosos estudios en los que se ha investigado la colaboración a partir de los datos cuantitativos de las intervenciones que permiten tener una visión muy general de las cantidades y flujos de las interacciones, sin entrar en el contenido de la interacción y las consecuencias respecto al aprendizaje.

Algunos trabajos que analizan el contenido de la interacción y utilizan el modelo de Garrison y Anderson (2005), como los presentados en el Capítulo 4, abordan el estudio de las contribuciones que las discusiones en línea puedan hacer en la promoción de estrategias de pensamiento crítico. Aún así, se concentran en categorizar las intervenciones y en el aspecto cuantitativo resultante.

Con el objeto de dar una solución se propone aplicar dos técnicas de minería de datos para extraer conocimiento de los foros de aprendizaje de las propuestas de e-learning que permitan establecer relaciones de ocurrencia de los indicadores, evaluar el nivel de presencia cognitiva en los foros del curso e identificar las categorías que brindan mayor información respecto a este nivel.

6.2. APORTES AL LOGRO DEL OBJETIVO

Construcción del instrumento para la categorización de la Dimensión Cognitiva

Para dotar de precisión al procedimiento de codificación en las subcategorías de análisis de Presencia Cognitiva, se construyó un instrumento con más detalle en la definición de cada indicador, con un alto grado de exclusividad y exhaustividad entre los mismos, que se incluye en el Anexo B.

Desarrollo de heurísticas específicas para categorizar los mensajes de los foros.

En este trabajo la unidad de análisis utilizada es el *mensaje completo* y se considera a éste cómo *el comprobante de la interacción*. Así como el comprobante de una compra-venta es la factura o el ticket en los cuales se detallan las cantidades y precios de los artículos involucrados en la transacción, en el mensaje se manifiestan los indicadores de las categorías social, docente y cognitiva involucradas en la interacción.

Cada mensaje se divide en partes que pueden ser identificadas en una de las categorías de análisis establecido. Un fragmento es una contribución interpretable y codificable en una de las categorías del sistema de análisis establecido (Coll Salvador, Bustos Sánchez, y Engel Rocamora, 2011). El mensaje completo puede estar formado por uno o varios fragmentos. La

identificación de los fragmentos y su codificación se hace por separado para cada una de las tres dimensiones. En lo que concierne al procedimiento de codificación, un mismo fragmento sólo puede ser codificado una vez en una categoría de cada dimensión. Los distintos fragmentos de un mensaje pueden ser codificados, si procede, en las distintas categorías de una misma dimensión.

En primer lugar se compilan los mensajes de los foros de cada grupo y se asigna un número correlativo a cada uno de ellos. Posteriormente, se codifican los mensajes en las categorías de Presencia Cognitiva, según el instrumento construido. La codificación se realiza de manera independiente por parte de tres analistas sobre los mismos textos. La revisión de la codificación tiene lugar de manera compartida entre los investigadores que discuten sus dificultades para asignar cada fragmento a alguna categoría. En esta línea, durante las sesiones convocadas se utiliza un control de fiabilidad por consenso entre los tres investigadores, donde los registros se comparan para controlar el nivel de acuerdo existente.

Diseño del instrumento para la recopilación de datos para reglas de asociación

Las reglas de asociación son una manera muy popular de expresar patrones de un conjunto de datos y surgieron inicialmente para afrontar el análisis de las cestas de la compra de los comercios. Estas cestas de compra se ven reflejadas en los comprobantes que se generan en cada una de las transacciones de compra-venta. En forma análoga a la aplicada a las cestas de compras se trabaja con esta técnica para extraer conocimiento de los foros de aprendizaje.

Para recopilar los datos a analizar con las reglas de asociación se establece que los *mensajes* de los foros se pueden expresar en una tabla, las *filas* de la tabla se refieren a los *mensajes* y las *columnas* son cada uno de *indicadores* de presencia cognitiva definidos. La tabla solo contiene valores binarios. Un 1 en la posición (i,j) indica que el mensaje i contiene el indicador j , mientras que un 0 indica que el mensaje no ha mostrado el indicador. En el Anexo B se muestran los instrumentos en forma detallada.

Diseño del instrumento para la recopilación de datos para árboles de decisión

Un árbol de decisión es un conjunto de condiciones organizadas en una estructura jerárquica, de tal manera que la decisión final a tomar se puede determinar siguiendo las condiciones que se cumplen desde la raíz del árbol hasta alguna de sus hojas.

Para evaluar el nivel de presencia cognitiva, los valores de los indicadores observados en los foros de aprendizaje se organizan en una tabla. En este caso, cada fila representa *un estudiante* que participa en el foros, las columnas son los indicadores de presencia cognitiva a alcanzar *para un objetivo de aprendizaje*. La tabla sólo contiene valores binarios. Un 1 en la posición (i,j) indica que el estudiante i ha alcanzado el indicador j , mientras que un 0 indica que el estudiante no ha logrado el indicador.

Para completar la tabla resta definir un resultado o “función objetivo” que se obtiene del conjunto de valores observados, de acuerdo con un dominio de conocimiento que se quiere adquirir. Este resultado para el dominio que se esta analizando es el “Nivel de Presencia Cognitiva” (NPC) alcanzado por el estudiante para un objetivo de aprendizaje, y tendrá los valores: Muy Alto, Alto, Medio o Bajo. (Anexo B)

6.3. CONCLUSIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN

Se ha logrado extraer conocimiento de los foros de aprendizaje de dos cursos de e-learning. El primero denominado *Enseñar y Aprender en el Aula Virtual* y el segundo *Prácticas Educativas en Entornos Virtuales*.

Se han descubierto reglas de asociación entre los indicadores de presencia cognitiva de un foro de aprendizaje del curso *Enseñar y Aprender en el Aula Virtual*.

Se construyó un árbol de decisión que permite evaluar el nivel de presencia cognitiva en un foro de aprendizaje del curso *Prácticas Educativas en Entornos Virtuales* e identificar las categorías que brindan mayor información respecto al nivel de presencia cognitiva.

Se considera que las técnicas de minería de datos tienen gran utilidad desde tres posibles campos de aplicación:

- en primer lugar, las reglas generadas permiten evaluar un foro para determinar el nivel de profundidad en términos de desarrollo de pensamiento crítico que se observa en una discusión en línea;
- la segunda aplicación se ubica en el nivel de diseño de un foro, para planificar el tipo de preguntas que se deben plantear de forma que se propicie un alto nivel de complejidad en la respuesta, y
- finalmente, una tercera aplicación se enfoca en la participación del tutor, quien puede utilizar los resultados observados para reorientar el desarrollo de la discusión; ya que las reglas ofrecen a los docentes un medio para evaluar rápidamente el carácter del discurso e informar si se requiere una intervención directa para hacerlo avanzar.

De esta manera con el conocimiento extraído de los foros de aprendizaje de los cursos de e-learning es posible formular estrategias preventivas y correctivas que permitan la mejora continua del proceso de enseñanza-aprendizaje.

Con el objetivo de seguir avanzando en la temática, se proponen a continuación algunos de los trabajos de investigación que podrían realizarse en el futuro:

- Enriquecer los instrumentos construidos incorporando los indicadores de la Presencia Docente y de Presencia Social y aplicarlos en la categorización de los mensajes.
- Aplicar técnicas minería de datos que permitan descubrir conocimiento de las interacciones de todas las presencias en los foros de aprendizaje.
- Desarrollar herramientas de software capaces de realizar análisis de las transacciones del e-learning en forma automática.

Por lo tanto, el reto consiste en proponer actividades que capitalicen las características y el potencial del medio en uso. Ello significa pensar de modo diferente sobre lo que puede ser una experiencia de e-learning. Implica reconocer la posibilidad de crear una comunidad de investigación cooperativa, crítica y en permanente mejoramiento.

CAPÍTULO 7

Bibliografía

BIBLIOGRAFÍA

Aballay, A., Carestia, N. y Martín, A (2007). *Aplicación de la Gestión de la Calidad Total a los Sistemas de Formación No-Presencial*. Ponencia presentada en EDUTEC 2007 Inclusión Digital en la Educación Superior. Desafíos y oportunidades en la Sociedad de la Información. Buenos Aires, Argentina. Recuperado el 18/07/2012 de <http://www.utn.edu.ar/aprobedutec07/docs/159.pdf>

Agrawal, Rakesh; Imielinski, Tomasz; Swami, Arun (1993) Mining association rules between sets of items in large databases. En *ACM SIGMOD Conference* (pp. 207-216) Washington DC, USA: ACM.

Agrawal, R., Srikant, R. (1994). *Fast Algorithms for Mining Association Rules*. Ponencia presentada en 20th International Conference on Very Large Databases. Santiago de Chile.

Anderson, T. (2001). The hidden curriculum of distance education. *Change Magazine*, 33(6), pp. 29-35.

Anderson, T., Rourke, L., Garrison, D.R. and Archer, W. (2001). Assessing teaching presence in a computer conferencing context. *Journal of the Asynchronous Learning Networks*, 5 (2). Recuperado el 15 de Agosto de 2013 de http://www.aln.org/alnweb/journal/Vol5_issue2/Anderson/5-2%20JALN%20Anderson%20Assessing.htm.

Ander-Egg, E. (2000). *Métodos y técnicas de investigación social. Cómo organizar el trabajo de investigación*. Argentina: Editorial Lumen Hvmanitas.

Arroyo Ávila, R., Reyes López, G. y García Bencomo, M. (2010). *Uso de discusiones virtuales en el desarrollo de habilidades cognitivas de estudiantes en la modalidad en línea de la Universidad Autónoma de Chihuahua*. Ponencia presentada en XII Asamblea General de ALAFEC. Lima, Perú. Recuperado el 5 de Abril de 2013 de <http://www.alafec.unam.mx/mem/lima/educacion/EDUC3.pdf>

Bussone, M.C. (2006). *Una experiencia de comunidad de indagación: hacia la comprensión de las interacciones realizadas en una comunidad virtual de aprendizaje.* Recuperado el 8 de Marzo de 2013 de <http://www.salvador.edu.ar/vrid/publicaciones/PonenciaBUSSONE.pdf>

Coll Salvador, C., Bustos Sánchez; A. y Engel Rocamora, A. (2011). Perfiles de participación y presencia docente distribuida en redes asíncronas de aprendizaje: la articulación del análisis estructural y de contenido. *Revista de Educación*, 354, pp. 657-688. Recuperado el 18 de Mayo de 2013 de http://www.revistaeducacion.mec.es/re354/re354_26.pdf

Fallas Monge, I. (2008). *¿Está mi foro produciendo pensamiento crítico?* Recuperado el 23 de Marzo de 2013 de <http://www.uned.ac.cr/ece/images/documents/documents2010/Estamiforoproduciendopensamientocritico.pdf>

Fayyad, U. and Stolorz, P. (1997). Data mining and KDD: promise and challenge. *Future Generation Computer System*, 13(2-3), pp. 99-115.

Garrison D.R. (2001). Research based continuing studies: A transformational leadership model. *Canadian Journal of University Continuing Education*, 27(1), pp. 77-97.

Garrison D.R. y Archer W. (2000). *A transactional perspective on teaching and learning: A framework for adult and higher education.* Oxford, US: Pergamon.

Garrison D.R. y T. Anderson (2005). *El e-learning en el siglo XXI. Investigación y práctica.* Barcelona: Octaedro S.L.

Garrison D.R., T. Anderson y Archer, W. (2000). Critical inquiry in a text-based environment: Computer Conferencing in higher education. *The Internet and Higher Education*, 2 (2-3), pp. 87-105.

Garrison D.R., T. Anderson y W. Archer (2001). Critical thinking, cognitive presence and computer conferencing in distance education, *Américal Journal of distance education*, **15** (1), pp. 7-23.

Gewerc B., A. (2008). ¿Qué estamos entendiendo por e-learning?. En A. Gewerc B. (Coord.). *Universidad y sociedad del conocimiento: ¿Es el e-learning la única respuesta?* (pp. 18-30) España: Universidad de Santiago de Compostela. Recuperado el 20 de Marzo de 2013 de http://unisc.usc.es/informes/Informe2007_2008.pdf

Gilbert, S.W. (2000). *So, why bother?* AAHESGIT, 49. Recuperado el 13 de Mayo de 2013 de <http://www.Tltgroup.org/gilbert/WhyBother.htm>

Gros Salvat, Begoña; Silva, Juan (2006). El problema del análisis de las discusiones asincrónicas en el aprendizaje colaborativo mediado. *RED. Revista de Educación a Distancia*, *V(16)*. Recuperado el 28 de Marzo de 2013 de <http://www.um.es/ead/red/16/gros.pdf>

Gunawardena, C., & McIsaac, M. (2004). Distance education. En D.H. Jonassen, *Handbook of research on educational communications and technology* (2da. Ed.) (pp. 355-395). New Jersey, EE.UU.: Lawrence Erlbaum Associates, Inc.

Hernández Orallo, J.; Ramírez Quintana, M.J.; Ferri Ramírez, C. (2004). *Introducción a la Minería de Datos*. España: Pearson Educación, S.A.

Hernández Sampieri, Roberto; Fernández Collado, Carlos; Baptista Lucio, Pilar (2008). Capítulo 4: Casos de Estudio [CD-ROM]. En R. Hernández Sampieri, C. Fernández Collado y P. Baptista Lucio, *Metodología de la Investigación* (4ta. Ed.). México: Mc. Graw Hill.

Hipp, J, Güntzer, U. y Nakhaeizaeh G. (2000). Algorithms for Association Rule Mining. A General Survey and Comparison. En: *ACM SIGKDD*, Boston: ACM.

Lipman, M. (1991). *Thinking in education*. Cambridge: Cambridge University Press.

Mertens, D.M. (2005), *Research and evaluation in Education and Psychology: integrating diversity with quantitative, qualitative, and mixed methods*. (2da. ed.). Thousand Oaks: Sage.

Meyer, K. (2004). Evaluating Online Discussions: Four Difference Frames of Analysis. *Journal of Asynchronous Learning Networks*, 8(2), pp. 101-114. Recuperado el 11 de Mayo de 2013 de http://www.sloan-c.org/publications/jaln/v8n2/pdf/v8n2_meyer.pdf.

Paul, Richard (1990). *Critical Thinking*. Rohnert Park. Sonoma State University. Recuperado el 11 de mayo de 2013 http://assets00.grou.ps/0F2E3C/wysiwyg_files/FilesModule/criticalthinkingandwriting/20090921185639-uxlhmlnvedpammxrz/CritThink1.pdf.

Perkins, C., & Murphy, E. (2006). Identifying and measuring individual engagement in critical thinking in online discussions: An exploratory case study. *Educational Technology & Society*, 9 (1), pp. 298-307.

Quinlan, JR (1986). Induction of decision trees. *Machine Learning*, 1(1), pp. 81–106.

Ramsden, P. (1988). Context and strategy: Situational influences on learning. En R.R. Schmeck (ed.) *Learning strategies and learning styles* (pp. 159-184). New York: Olenum.

Ripa, María Estela (2007). ¿Cómo abordar el estudio de una comunidad de aprendizaje Blended Learning. Rodríguez Illera, José Luis. (Coord.) Comunidades virtuales de práctica y de aprendizaje. *Revista Electrónica de la Educación: Educación y Cultura en la Sociedad de la Información*, 8(3). Recuperado el 5 de Febrero de 2013 de http://www.usal.es/~teoriaeducacion/rev_numero_08_03/n8_03_ripa

Rourke, L., Anderson, T., Archer, W. and Garrison, D.R. (1999). Assessing social presence in asynchronous, text based computer conference. *Journal of Distance Education*, 14(3), pp. 51-70.

Rourke, L. and Anderson, T. (2001). *Running Head: Peer teams leading online discussion. Using Peer Teams to Lead Online Discussions.* Recuperado el 10 de Marzo de 2013 de www.communitiesofinquiry.com/documents/peerteams.pdf

Schunk, D.H. (2000). *Learning Theories* (3° Ed). Englewood Cliff, New Jersey: Prentice Hall.

Srivastava, J., Mobasher, B. y Cooley, R. (2000). Automatic Personalization Based on Web Usage Mining. *Communications of the Association of Computing Machinery.* pp. 142-151.

Stake, R. (2000). Case Studies. En N.Denzin, N. y Y. Lincoln (Eds.). *Handbook of qualitative research* (pp. 236-247) (2° Ed.). Thousand Oaks: sage Publications.

Waisman, E.; Gómez, C.; Olivares, M. y Font, G. (2011). *Propuesta pedagógica.* Recuperado el 08 de Marzo de 2012 de http://www.ecampus.unsj.edu.ar/file.php/106/Propuesta_Pedagogica.pdf

Williams, M., Unrau, Y.A. y Grinnell, P.M. (2005) The qualitative research approach. En R.M. Grinnell y Y.A.Unrau (Eds.). *Social work: Research and evaluation. Quantitative and qualitative approaches* (pp. 75-87). (7° Ed.). Nueva York: Oxford University Press.

Witten, I.H. and Frank E. (1999). *Data Mining. Practical Machine Learning Tools and Techniques with Java Implementations.* California: Morgan Kaufmann.

Yin, R.K. (2003). *Case study research. Design and methods* (3° Ed.). California: Sage.

Zytkow,, J y Klosgen, W. (2001). *Handbook of Data Mining and Knowledge Discovery.* USA: Oxford University Press.

ANEXO A

WEKA

ANEXO A: WEKA

A.1. INTRODUCCIÓN

La Weka (*Gallirallus australis*) es un ave endémica de Nueva Zelanda. Este ave da nombre a una extensa colección de algoritmos de conocimiento de máquina desarrollados por la Universidad de Waikato (Nueva Zelanda) e implementados en Java; útiles para ser aplicados sobre datos mediante las interfaces que ofrece o para embeberlos dentro de cualquier aplicación. Además Weka contiene las herramientas necesarias para realizar transformaciones sobre los datos, tareas de clasificación, regresión, clustering, asociación y visualización. Weka está diseñado como una herramienta orientada a la extensibilidad por lo que añadir nuevas funcionalidades es una tarea sencilla.

La licencia de Weka es GPL, lo que significa que este programa es de libre distribución y difusión. Además, ya que Weka está programado en Java, es independiente de la arquitectura, pues funciona en cualquier plataforma sobre la que haya una máquina virtual Java disponible. La página principal correspondiente a este software se encuentra en la dirección: <http://www.-cs.waikato.ac.nz/ml/weka/>

A.2. ARCHIVOS EN WEKA

Weka trabaja principalmente sobre archivos un con formato denominado arff , acrónimo de *Attribute-Relation File Format*. Este formato está compuesto por una estructura claramente diferenciada en tres partes:

Cabecera. En esta sección se define el nombre de la relación. Su sintaxis es la siguiente:

@relation <nombre-de-la-relación>

Donde <nombre-de-la-relación> es de tipo String.

Ejemplo: @relation presenciacognitiva

Si dicho nombre contiene algún espacio será necesario expresarlo entrecomillado.

Declaraciones de atributos. En esta sección se declaran los atributos que compondrán el archivo junto a su tipo, su sintaxis es la siguiente:

@attribute <nombre-del-atributo> <tipo>

donde <nombre-del-atributo> es de tipo String y tiene las mismas restricciones que el caso anterior.

Weka acepta diversos tipos, estos son:

a) **NUMERIC** Expresa números reales- la separación de la parte decimal y entera de los números reales se realiza mediante un punto en vez de una coma.

b) **INTEGER** Expresa números enteros.

c) **DATE** Expresa fechas, para ello este tipo debe ir precedido de una etiqueta de formato entrecomillada.

La etiqueta de formato está compuesta por caracteres separadores (guiones y/o espacios) y unidades de tiempo:

dd Día.

MM Mes.

yyyy Año.

HH Horas.

mm Minutos.

ss Segundos.

d) **STRING** Expresa cadenas de texto, con las restricciones del tipo String comentadas anteriormente.

e) **ENUMERADO** El identificador de este tipo consiste en expresar entre llaves y separados por comas los posibles valores (caracteres o cadenas de caracteres) que puede tomar el atributo. Por ejemplo, el atributo que indica nivel de presencia cognitiva, podría definirse:

@attribute nivelpresenciacognitiva {muyalto, alto, medio, bajo}

Sección de datos. Se declaran los datos que componen cada transacción separando entre comas los atributos y con saltos de línea las relaciones. Ejemplo:

@data

0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,bajo

1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,bajo

En el caso de que algún dato sea desconocido, este se expresará con un símbolo de cerrar interrogación (“?”).

Es posible añadir comentarios con el símbolo “%”, que indicará que desde ese símbolo hasta el final de la línea es todo un comentario. Los comentarios pueden situarse en cualquier lugar del fichero.

A.3. FICHEROS PARA NIVEL DE PRESENCIA COGNITIVA

Fichero de entrenamiento

@relation presenciacognitiva

@attribute hdsensaciondeperplejidad {0,1}

@attribute hdreconocerelproblema {0,1}

@attribute hdconfusion {0,1}

@attribute hdotros {0,1}

@attribute exdivergencia {0,1}

@attribute exintercambiodeinformacion {0,1}

@attribute exsugerencias {0,1}

@attribute exlluviadeideas {0,1}

@attribute exsaltosintuitivos {0,1}

@attribute inconvergencia {0,1}

@attribute insintesis {0,1}

@attribute insoluciones {0,1}

@attribute reaplicar {0,1}

@attribute recomprobar {0,1}

@attribute redefender {0,1}

@attribute nivelpresenciacognitiva {muyalto,alto, medio,bajo}

@data

0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,bajo

1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,bajo

0,1,0,0,0,0,0,0,0,0,0,0,0,0,bajo
 0,0,1,0,0,0,0,0,0,0,0,0,0,0,bajo
 0,0,0,1,0,0,0,0,0,0,0,0,0,0,bajo
 0,0,0,0,1,0,0,0,0,0,0,0,0,0,bajo
 0,0,0,0,0,1,0,0,0,0,0,0,0,0,bajo
 0,1,0,0,0,1,0,0,0,0,0,0,0,0,bajo
 0,1,0,0,0,0,1,0,0,0,0,0,0,0,bajo
 0,1,0,0,0,0,0,1,0,1,0,0,0,0,medio
 0,1,0,0,0,1,0,0,0,1,0,0,0,0,medio
 0,1,0,0,0,0,0,0,0,1,0,0,0,0,medio
 0,1,0,0,0,1,1,0,0,1,0,0,0,0,medio
 0,1,0,0,0,1,0,1,0,1,0,0,0,0,medio
 0,1,0,0,0,1,0,0,0,1,1,0,0,0,0,alto
 0,1,0,0,0,1,1,0,0,0,1,1,0,0,0,alto
 0,1,0,0,0,1,0,0,0,1,1,0,0,0,0,alto
 1,1,0,0,0,1,0,0,0,0,1,1,0,0,0,alto
 1,1,0,0,0,1,0,0,0,1,1,1,0,0,0,alto
 0,1,0,0,0,1,0,0,0,0,1,1,1,0,0,muyalto
 1,1,0,0,0,1,0,0,0,0,1,1,1,1,0,muyalto
 1,1,0,0,0,1,0,0,0,0,1,1,1,1,1,muyalto
 1,1,0,0,0,1,1,1,1,1,1,1,1,1,muyalto

Fichero de prueba

@relation presenciacognitiva

@attribute hdsensaciondeperplejidad {0,1}

@attribute hdreconocerelproblema {0,1}

@attribute hdconfusion {0,1}

@attribute hdotros {0,1}

@attribute exdivergencia {0,1}

@attribute exintercambiodeinformacion {0,1}

@attribute exsugerencias {0,1}

@attribute exlluviadeideas {0,1}

@attribute exsaltosintuitivos {0,1}
 @attribute inconvergencia {0,1}
 @attribute insintesis {0,1}
 @attribute insoluciones {0,1}
 @attribute reaplicar {0,1}
 @attribute recomprobar {0,1}
 @attribute redefender {0,1}
 @attribute nivelpresenciacognitiva {muyalto,alto,medio,bajo}

@data

0,0,0,1,0,0,0,1,0,1,1,0,0,0,0,alto
 0,0,0,1,0,0,0,0,0,1,0,0,0,0,0,medio
 0,1,0,1,0,1,0,1,0,1,0,0,0,0,0,medio
 0,1,0,1,0,1,0,0,0,0,0,0,0,0,0,bajo
 0,0,0,0,0,1,0,1,0,1,0,0,0,0,0,medio
 0,0,0,0,0,1,0,0,0,1,0,0,0,0,0,medio
 0,0,0,0,0,1,1,1,0,1,0,0,0,0,0,medio
 0,0,0,0,0,1,1,1,0,1,1,0,0,0,0,alto

A.4. EJECUCIÓN DE WEKA

Una vez que Weka está instalado y en ejecución, aparecerá una ventana denominada *selector de interfaces*, que permite seleccionar la interfaz con la se comienza a trabajar. Las posibles interfaces a seleccionar son *Simple Cli*, *Explorer*, *Experimenter* y *Knowledge flow* como muestra la figura A.1.

Simple Cli : Simple CLI es una abreviación de *Simple Client*. Esta interfaz proporciona una consola para poder introducir mandatos. A pesar de ser en apariencia muy simple, como se observa en la figura A.2 es extremadamente potente porque permite realizar cualquier operación soportada por Weka de forma directa; no obstante, es muy complicada de manejar ya que es necesario un conocimiento completo de la aplicación. Su utilidad es pequeña desde que se fue recubriendo Weka con interfaces.

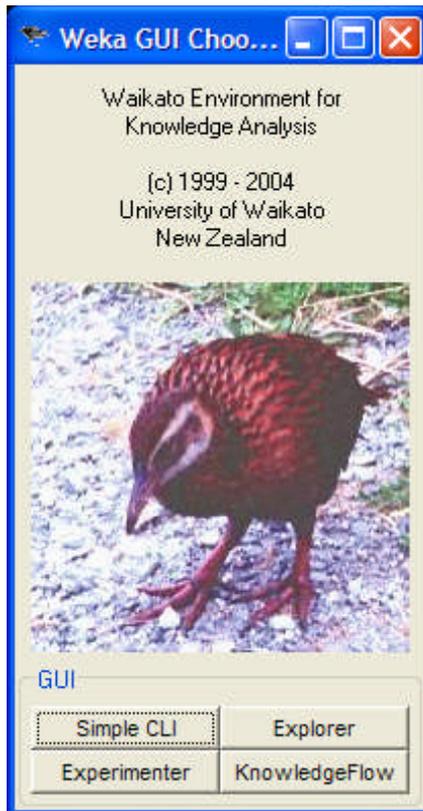


Figura A.1. Ventana de selección de Interfaz

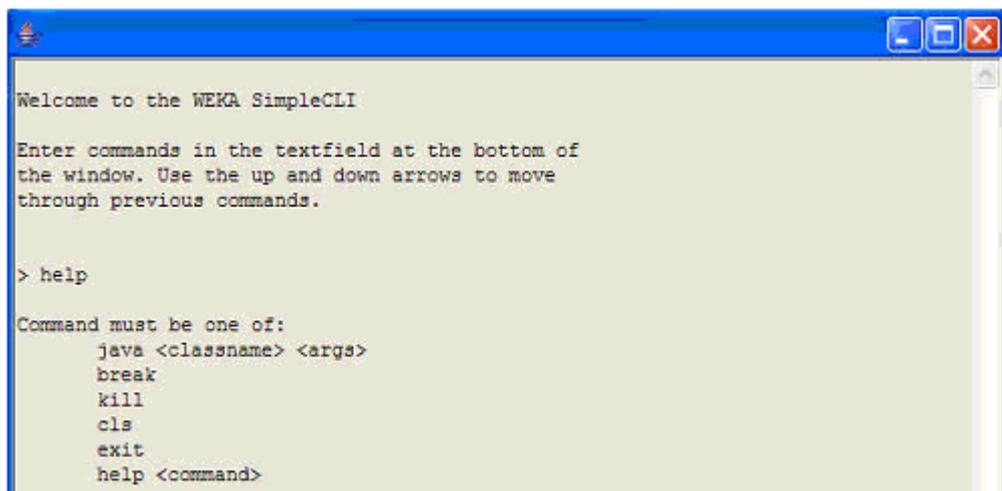


Figura A.2. Ventana principal Simple CLI

Explorer: El modo *Explorador* es el modo más usado y más descriptivo. Éste permite realizar operaciones sobre un sólo archivo de datos. Se muestra en figura A.3.

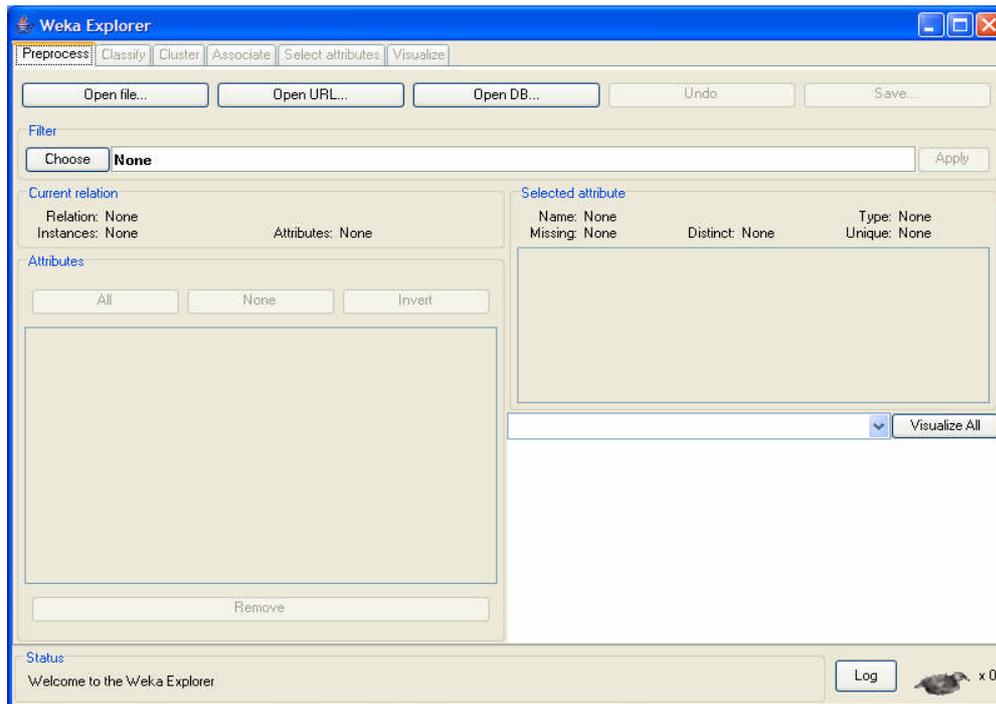


Figura A.3. Ventana principal Explorer

Experimenter: El modo experimentador es un modo muy útil para aplicar uno o varios métodos de clasificación sobre un gran conjunto de datos y, luego poder realizar contrastes estadísticos entre ellos y obtener otros índices estadísticos. Ver en figura A.4.

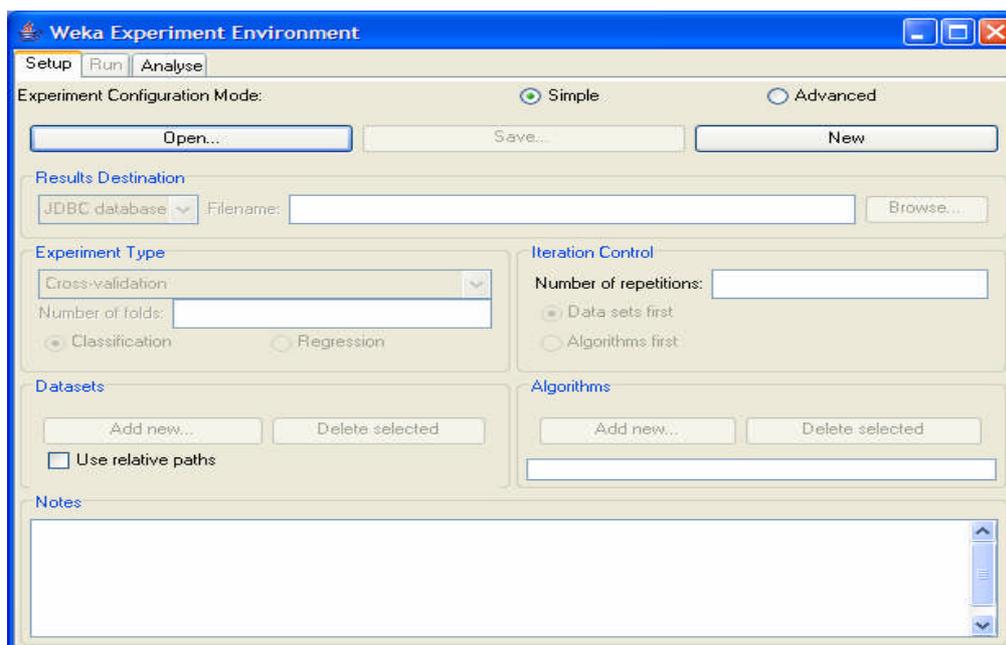


Figura A.4. Ventana principal Experimenter

Knowledge flow: Esta última interface de Weka es quizá la más cuidada y la que muestra de una forma más explícita el funcionamiento interno del programa. Su funcionamiento es gráfico y se basa en situar en el panel de trabajo, elementos base que están situados en la parte superior de la ventana de manera que se cree un *circuito* que defina el experimento. Se observa en la figura A.5.

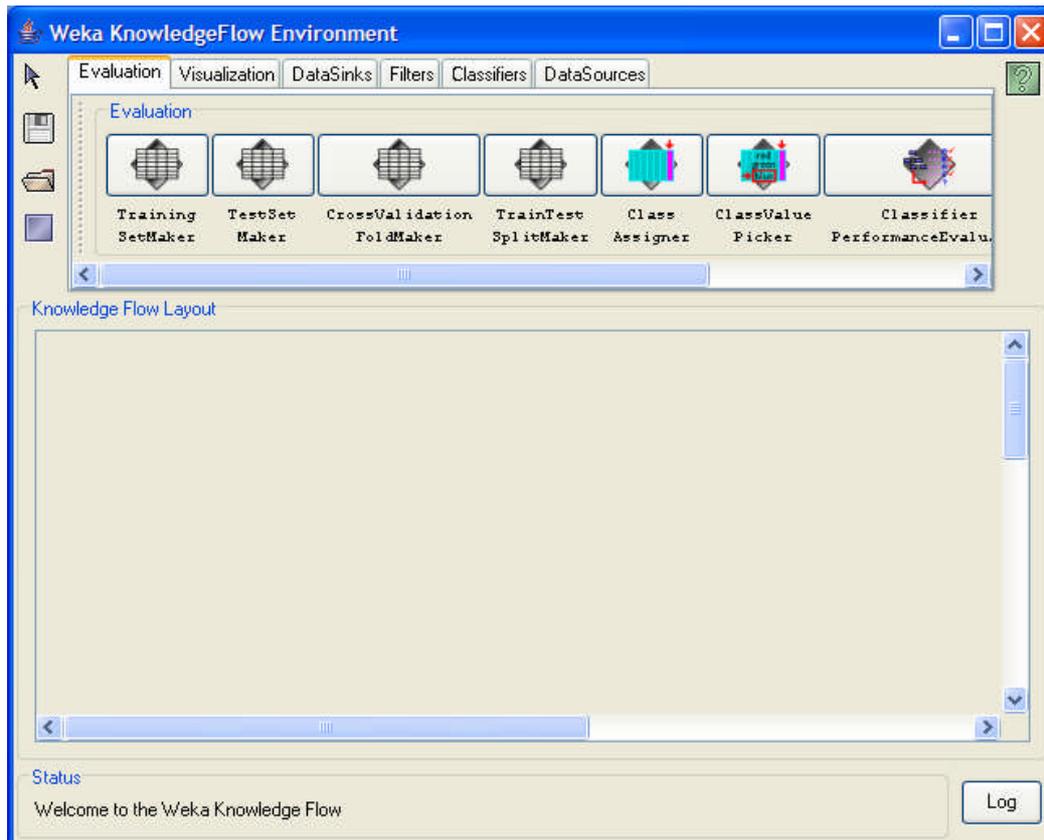


Figura A.5. Ventana principal Knowledge flow

De las opciones provistas por Weka en el selector de interfaces, se trabaja con el Explorador.

A.5. MODO EXPLORER

El explorador permite tareas de:

- 1 Preprocesado de los datos. y aplicación de filtros.
- 2 Clasificación.
- 3 Clustering.
- 4 Búsqueda de Asociaciones.

- 5 Selección de atributos.
- 6 Visualización de datos.

A continuación, se describe la utilización de Preprocesado de datos, Clasificación y Búsqueda de asociaciones.

A.5.1. Preprocesado de datos

El primer paso para comenzar a trabajar con el explorador es definir el origen de los datos. Por medio de la opción Open File, puede seleccionarse un archivo. Aunque el formato por defecto de Weka es el arff eso no significa que sea el único que admita, para ello tiene intérpretores de otros formatos, ya que soporta diferentes fuentes –arfff, C4.5, Csv, base de datos, etc.

Para cargar un archivo simplemente debe buscarse la ruta donde este se encuentra y seleccionarlo.

Una vez seleccionado el origen de los datos puede aplicarse algún filtro sobre él o bien pasar a las siguientes secciones y realizar otras tareas.

Por medio de la opción Open File, se abre el archivo presenciacognitiva.arff. Este archivo contiene 16 atributos que son los indicadores de presencia cognitiva definidos en el tabla B.1. Estos indicadores se transforman en los atributos de la relación presencia cognitiva, de la siguiente forma:

Para la categoría Hecho Desencadenante se definen cuatro atributos:

```
@attribute hdsensaciondeperplejidad {0,1}
@attribute hdreconocerelproblema {0,1}
@attribute hdconfusion {0,1}
@attribute hdotros {0,1}
```

Para la categoría Exploración se definen cinco atributos:

@attribute exdivergencia {0,1}
@attribute exintercambiodeinformacion {0,1}
@attribute exsugerencias {0,1}
@attribute exlluviadeideas {0,1}
@attribute exsaltosintuitivos {0,1}

Para la categoría Integración se definen tres atributos:

@attribute inconvergencia {0,1}
@attribute insintesis {0,1}
@attribute insoluciones {0,1}
@attribute reaplicar {0,1}
@attribute recomprobar {0,1}
@attribute redefender {0,1}

El archivo presenciacognitiva.arff mostrado en el inciso A.3 se genera a partir de los valores de los indicadores de presencia cognitiva observados en los foros de aprendizaje. Cada instancia representa los 15 valores que se asignan a los atributos para *un estudiante* que participa en los foros de aprendizaje. La instancia sólo contiene valores binarios. Un 1 en la posición j indica que el estudiante ha alcanzado el indicador j , mientras que un 0 indica que el estudiante no ha logrado el indicador.

Para completar la tabla resta definir un resultado o *función objetivo* que se obtiene del conjunto de valores observados, de acuerdo con un dominio de conocimiento que se quiere adquirir. Este resultado para el dominio que se analiza es el atributo *Nivel de Presencia Cognitiva* alcanzado por el estudiante para un objetivo de aprendizaje, y tendrá los valores: *Muy Alto, Alto, Medio o Bajo*.

@attribute nivelpresenciacognitiva {muyalto,alto,medio,bajo}

En un ambiente de enseñanza en línea, el docente debe planificar las actividades de aprendizaje y los objetivos que deberán alcanzar los estudiantes con cada una de ellas. Dichas actividades permitirán que el estudiante construya significados mediante la reflexión continua en una comunidad de aprendizaje, es decir logre alcanzar distintos niveles de presencia cognitiva.

En base a los objetivos, las actividades y los criterios de evaluación contenidos en la planificación de la práctica de aprendizaje grupal de la unidad I, el equipo docente construye

el conjunto de entrenamiento para la clasificación del nivel de presencia cognitiva, que se detalla en el inciso A.3.

La figura A.6. es la ventana que muestra la herramienta con cada uno de los atributos que componen los datos, junto con un resumen con estadísticas de los mismos, en este caso número de instancias distintas.

En el cuadrante inferior derecho aparece una representación gráfica del atributo seleccionado- *hdsensaciondeperplejidad* Encima de ésta hay un menú desplegable que permite variar el atributo de referencia, que se representará en color para contrastar ambos atributos. En este caso el atributo que acompaña es *nivelpresenciacognitiva*.

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Open file... | Open URL... | Open DB... | Undo | Save...

Filter: Choose None Apply

Current relation: Relation: presenciacognitiva, Instances: 23, Attributes: 16

Attributes: All | None | Invert

No.	Name
1	<input checked="" type="checkbox"/> hdsensaciondeperplejidad
2	<input type="checkbox"/> hdreconocerelproblema
3	<input type="checkbox"/> hdconfusion
4	<input type="checkbox"/> hdotros
5	<input type="checkbox"/> exdivergencia
6	<input type="checkbox"/> exintercambiodeinformacion
7	<input type="checkbox"/> exsugerencias
8	<input type="checkbox"/> exluviadeideas
9	<input type="checkbox"/> exsaltosintuitivos
10	<input type="checkbox"/> inconvergencia
11	<input type="checkbox"/> insintesis
12	<input type="checkbox"/> insoluciones
13	<input type="checkbox"/> reaplicar
14	<input type="checkbox"/> recomprobar

Remove

Status: OK

Log x 0

Selected attribute: Name: hdsensaciondeperplejidad, Type: Nominal, Missing: 0 (0%), Distinct: 2, Unique: 0 (0%)

Label	Count
0	17
1	6

Class: nivelpresenciacognitiva (Nom) Visualize All

Figura A.6 Ventana principal de preprocesado

La primera columna del cuadrante inferior derecho hace referencia al valor '0' del atributo **hdsensaciondeperplejidad** y la cantidad de transacciones en que este atributo ocurre en los datos, 17 en este caso.

La segunda columna del cuadrante inferior derecho hace referencia al valor '1' del atributo **hdsensaciondeperplejidad** y la cantidad de transacciones en que este atributo ocurre en los datos, 6 en este caso.

Los cuatro colores en los que esta columna se subdivide representan la cantidad de veces que el mencionado valor se encuentra acompañado por cada uno de los valores del atributo **nivelpresenciacognitiva**.

El valor '0' de **hdsensaciondeperplejidad** ocurre:

- 1 sola vez con el valor *muyalto* de **nivelpresenciacognitiva** (color azul)
- 3 veces con el valor *alto* de **nivelpresenciacognitiva** (color rojo)
- 5 veces con el valor *medio* de **nivelpresenciacognitiva** (color verde)
- 8 veces con el valor *bajo* de **nivelpresenciacognitiva** (color gris)

El valor '1' de **hdsensaciondeperplejidad** ocurre:

- 3 veces con el valor *muyalto* de **nivelpresenciacognitiva** (color azul)
- 2 veces con el valor *alto* de **nivelpresenciacognitiva** (color rojo)
- 1 sola vez con el valor *bajo* de **nivelpresenciacognitiva** (color gris)

Pulsando el botón **Visualize all** se abre una ventana desplegable como la de la figura A.7 que muestra todas las gráficas pertenecientes a todos los atributos, en relación con el atributo seleccionado.

Esta ventana presenta para cada uno de los 16 atributos, la cantidad de transacciones en las que cada valor del atributo coexiste con cada valor del atributo seleccionado en este caso **hdsensaciondeperplejidad**.

Se muestran 16 gráficos, el primero es el explicado anteriormente figura A.6, los otros, se interpretan a continuación los correspondientes a la categoría Hecho desencadenante y los indicadores de las demás categorías se interpretan en forma similar.

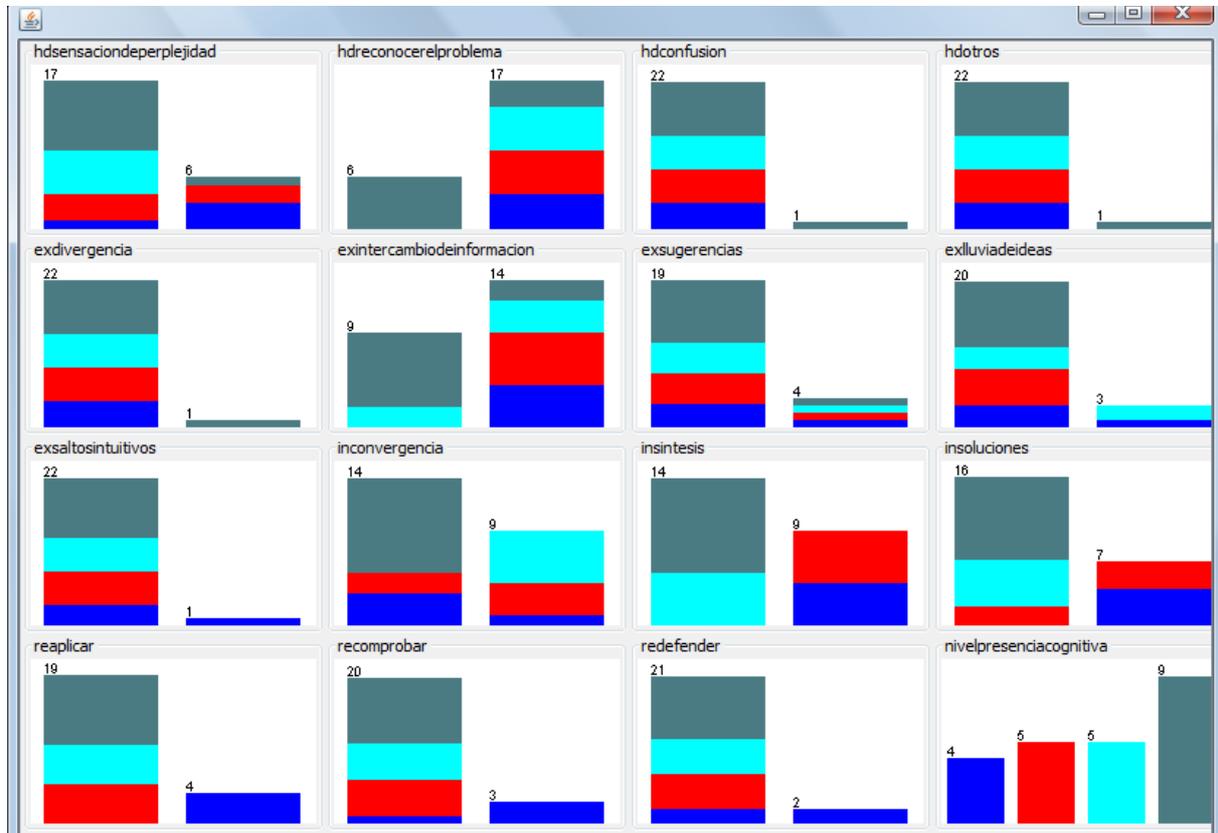


Figura A.7. Ventana de Visualize all

El valor '0' de **hdreconocerelproblema** ocurre:

6 veces con el valor *bajo* de **nivelpresenciacognitiva** (color gris)

El valor '1' de **hdreconocerelproblema** ocurre:

4 veces con el valor *muyalto* de **nivelpresenciacognitiva** (color azul)

5 veces con el valor *alto* de **nivelpresenciacognitiva** (color rojo)

5 veces con el valor *medio* de **nivelpresenciacognitiva** (color verde)

3 veces con el valor *bajo* de **nivelpresenciacognitiva** (color gris)

El valor '0' de **hdconfusion** ocurre:

4 veces con el valor *muyalto* de **nivelpresenciacognitiva** (color azul)

5 veces con el valor *alto* de **nivelpresenciacognitiva** (color rojo)

- 5 veces con el valor *medio* de **nivelpresenciacognitiva** (color verde)
- 8 veces con el valor *bajo* de **nivelpresenciacognitiva** (color gris)

El valor '1' de **hdconfusion** ocurre:

- 1 sola vez con el valor *bajo* de **nivelpresenciacognitiva** (color gris)

El valor '0' de **hdotros** ocurre:

- 4 veces con el valor *muyalto* de **nivelpresenciacognitiva** (color azul)
- 5 veces con el valor *alto* de **nivelpresenciacognitiva** (color rojo)
- 5 veces con el valor *medio* de **nivelpresenciacognitiva** (color verde)
- 8 veces con el valor *bajo* de **nivelpresenciacognitiva** (color gris)

El valor '1' de **hdotros** ocurre:

- 1 sola vez con el valor *bajo* de **nivelpresenciacognitiva** (color gris)

A.5.2. Clasificación

La segunda pestaña *Classify* permite aplicar métodos de clasificación la ventana se muestra en la figura A.8.

Se pretende construir un modelo que permita predecir la categoría de las instancias en función de una serie de atributos de entrada. En el caso de WEKA, la clase es simplemente uno de los atributos simbólicos disponibles, que se convierte en la variable objetivo a predecir. Por defecto, es el último atributo (última columna) a no ser que se indique otro explícitamente. La configuración de la clasificación se muestra en la figura A.8.

En la parte superior, se observa el botón 'Chose' que sirve para seleccionar el algoritmo de clasificación, ver en la figura A.9.

Se selecciona el algoritmo J48 de WEKA que es una implementación del algoritmo C4.5, uno de los algoritmos de minería de datos que más se ha utilizado en multitud de aplicaciones. No se describen todos los parámetros de su configuración, se explica sólo uno de los más impor-

tantes, el factor de confianza para la poda, *confidence level*, puesto que influye notoriamente en el tamaño y capacidad de predicción del árbol construido.

Una explicación simplificada de este parámetro de construcción del árbol es la siguiente: para cada operación de poda, define la probabilidad de error que se permite a la hipótesis de que el empeoramiento debido a esta operación sea significativo. Cuanto más baja se haga esa probabilidad, se exigirá que la diferencia en los errores de predicción antes y después de podar sea más significativa para no podar. El valor por defecto de este factor es del 25%, y conforme va bajando se permiten más operaciones de poda y por tanto llegar a árboles cada vez más pequeños. Otra forma de variar el tamaño del árbol es a través de un parámetro que especifica el mínimo número de instancias por nodo, si bien es menos elegante puesto que depende del número absoluto de instancias en el conjunto de partida.

Una vez elegido el clasificador el próximo paso es la configuración del modo de evaluación del clasificador. El resultado de aplicar el algoritmo de clasificación se efectúa comparando la clase predicha con la clase real de las instancias. Esta evaluación puede realizarse de diferentes modos, según la selección en el cuadro *Test options* como se observa en la figura A.10. Weka proporciona 4 modos de prueba: *Use training set*, *Supplied test set*, *Cross-validation*, *Percentage Split*.

- *Use training set*: esta opción evalúa el clasificador sobre el mismo conjunto sobre el que se construye el modelo predictivo para determinar el error, que en este caso se denomina *error de resustitución*. Por tanto, esta opción puede proporcionar una estimación demasiado optimista del comportamiento del clasificador, al evaluarlo sobre el mismo conjunto sobre el que se hizo el modelo.
- *Supplied test set*: evaluación sobre conjunto independiente. Esta opción permite cargar un conjunto nuevo de datos. Sobre cada dato se realizará una predicción de clase para contar los errores.
- *Cross-validation*: evaluación con validación cruzada. Esta opción es la más elaborada y costosa. Se realizan tantas evaluaciones como se indica en el parámetro *Folds*. Se dividen las instancias en tantas carpetas como indica este parámetro y en cada evaluación se toman las ins-

tancias de cada carpeta como datos de test, y el resto como datos de entrenamiento para Análisis de Datos en WEKA construir el modelo. Los errores calculados son el promedio de todas las ejecuciones.

- *Percentage split*: esta opción divide los datos en dos grupos, de acuerdo con el porcentaje indicado (%). El valor indicado es el porcentaje de instancias para construir el modelo, que a continuación es evaluado sobre las que se han dejado aparte. Cuando el número de instancias es suficientemente elevado, esta opción es suficiente para estimar con precisión las prestaciones del clasificador en el dominio.

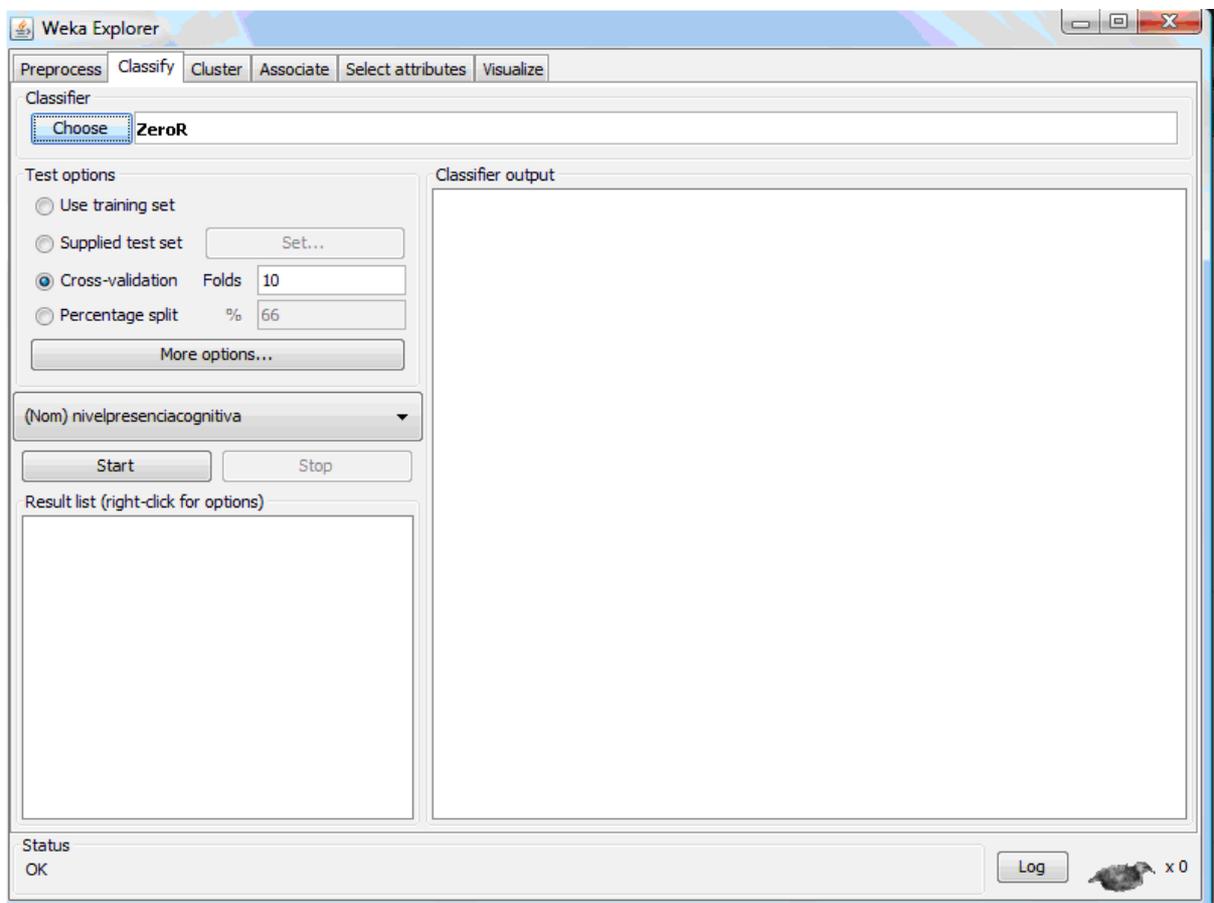


Figura A.8. Ventana principal de Classify

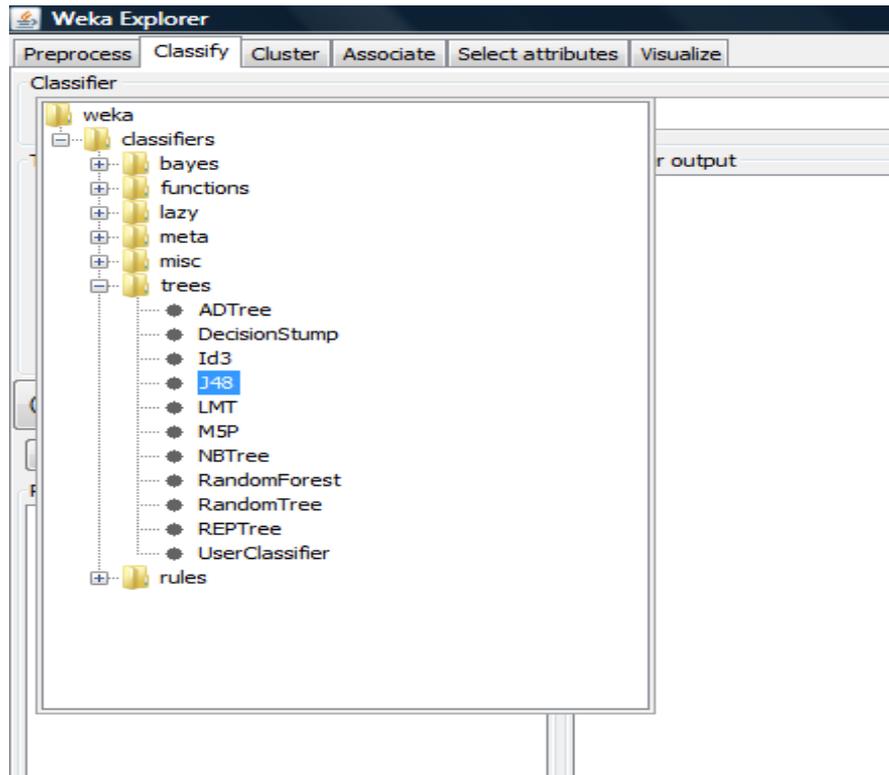


Figura A.9. Ventana que permite la selección de clasificadores

More Options abre un cuadro con opciones adicionales como muestra la figura A.11.:

Output model: permite visualizar en modo texto y, con algunos algoritmos, en modo gráfico el modelo construido por el clasificador, árbol, reglas, etc.

Output per-class stats: obtiene estadísticas de los errores de clasificación por cada uno de los valores que toma el atributo de clase

Output entropy evaluation measures: genera medidas de evaluación de entropía

Store predictions for visualization: permite analizar los errores de clasificación en una ventana de visualización

Cost-sensitive evaluation: con esta opción se puede especificar una función con costos relativos de los diferentes errores, que se rellena con el botón *Set*, en el ejemplo se utilizan los valores por defecto de estas últimas opciones.

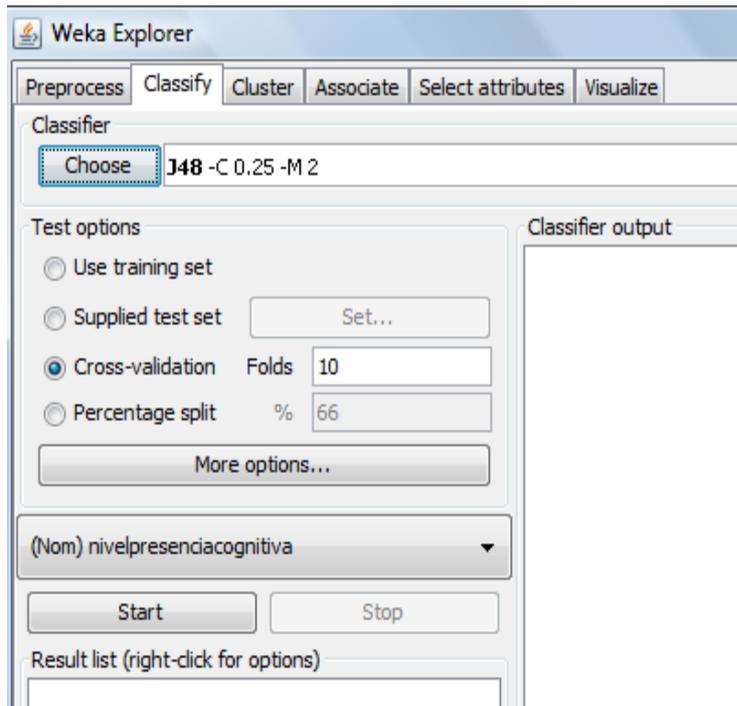


Figura A.10. Selección del Modo de evaluación

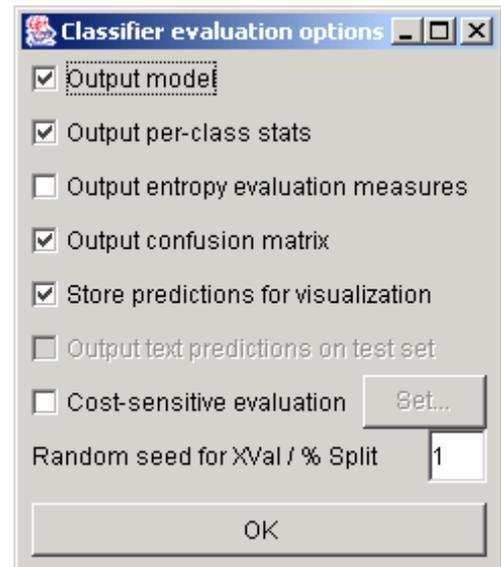


Figura A.11. Opciones adicionales

Una vez que se ejecuta el clasificador seleccionado, en la ventana de texto de la derecha aparece *classifiers output* o salida del clasificador que es un listado con los resultados, dónde se detalla: *run information*, *classifier model* y *evaluation*.

Run information o información de ejecución detalla el algoritmo utilizado (`weka.classifiers.trees.J48 -C 0, 25 -M 2`), el nombre de la relación (*presenciacognitiva*), la cantidad de instancias o filas procesadas (23), la cantidad de atributos (16), los nombres de cada atributo y el modo en que se ejecuta la clasificación (*evaluate on training data*) la imagen se muestra en la figura A.12.

Classifier model o modelo de clasificación indica el modo en que se ejecuta el clasificador *full training set* en este caso y detalla las reglas que permiten clasificar cada hoja del árbol. Es interesante observar que de los 16 atributos iniciales aparecen sólo 3 atributos en las reglas que genera. En el inciso 5.6.3 se explican estas reglas y la imagen se observa en la figura A.13.

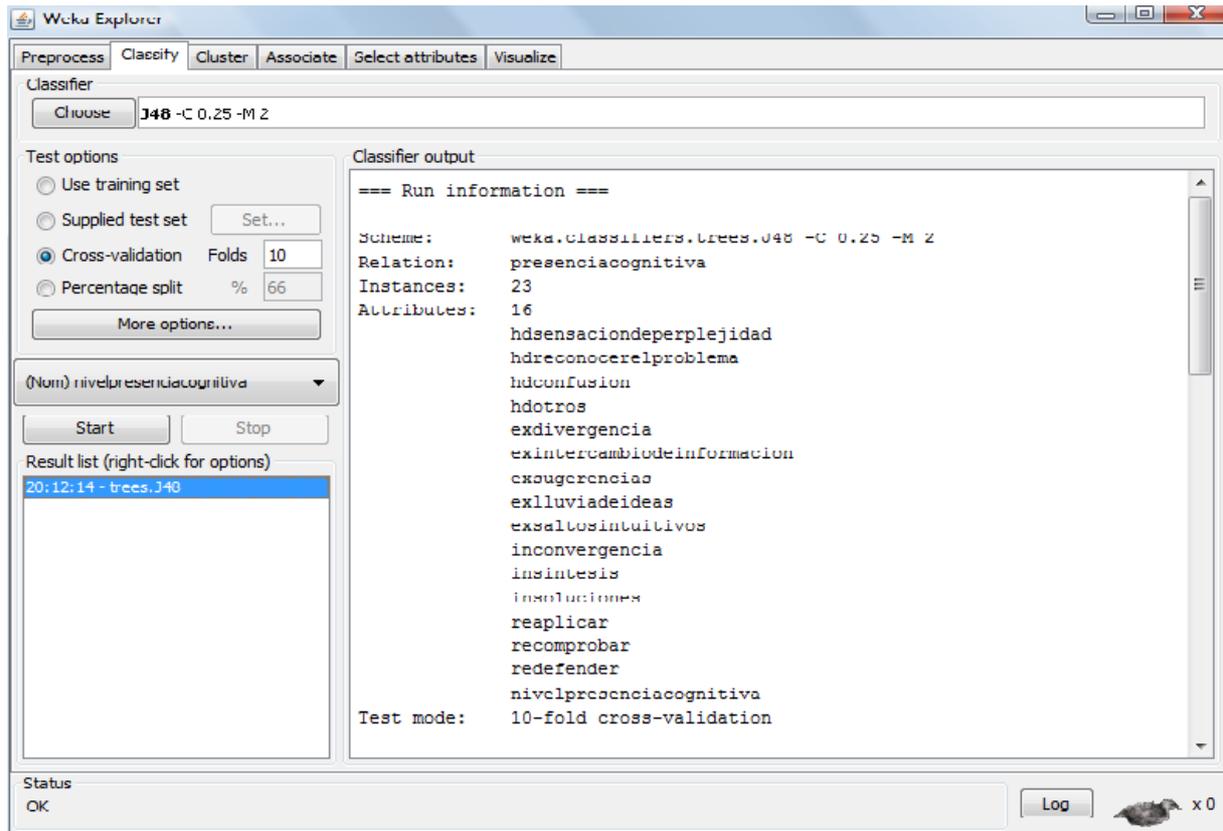


Figura A.12. Información de ejecución

```

=== Classifier model (full training set) ===

J48 pruned tree
-----

insintesis = 0
|  inconvergencia = 0: bajo (9.0)
|  inconvergencia = 1: medio (5.0)
insintesis = 1
|  reaplicar = 0: alto (5.0)
|  reaplicar = 1: muyalto (4.0)

Number of Leaves :    4
Size of the tree :    7

Time taken to build model: 0.05 seconds
  
```

Figura A.13. Modelo de clasificación

Con referencia a los resultados de evaluación del clasificador se pueden destacar tres tipos de informes *Summary*, *Detailed Accuracy by Class* y *Confusion Matrix*.

```

=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances      23          100    %
Incorrectly Classified Instances    0           0     %
Kappa statistic                    1
Mean absolute error                 0
Root mean squared error             0
Relative absolute error              0     %
Root relative squared error          0     %
Total Number of Instances          23

```

Figura A.14. Resumen

El primer informe *Summary* o resumen muestra el porcentaje global de errores cometidos en la evaluación. En la figura A.14 se observan los siguientes resultados: *Correctly Classified Instances*, *Incorrectly Classified Instances*, *Kappa statistic*, *Mean absolute error*, *Root mean squared error*, *Relative absolute error*, *Root relative squared error*, *Total Number of Instances*.

Correctly Classified Instances son las instancias correctamente clasificadas o modeladas, en este caso 23.

Incorrectly Classified Instances son las instancias incorrectamente clasificadas, en este caso 0.

Kappa Statistic o estadístico Kappa (coeficiente Kappa) se usa como un significado del grado de clasificación en datos categóricos. Un valor en el coeficiente Kappa de 1 significa un modelado estadísticamente perfecto, mientras un valor de 0 significa que cada valor del modelo fue diferente al valor actual o real. Un coeficiente kappa 0.7 o mayor es lo anhelado como correlación estadística Buena, pero por supuesto, el valor más alto da la mejor correlación:

Root mean-squared error o error absoluto medio es el promedio de la diferencia entre el valor predicho y el actual de testeo para todos los casos de testeo. Es el error de predicción promedio.

Root relative squared error o error cuadrático relativo es el total de los errores cuadráticos relativo a los errores cuadráticos respecto del valor medio absoluto. En este caso se exagera el error en aquellos casos en que el error de predicción es significativamente mayor que el error medio.

Relative absolute error o error absoluto relativo es el error absoluto total relativo al error que debería existir si la predicción se hubiera realizado como el promedio de los valores actuales.

Total Number of Instances es el número total de instancias modeladas en este caso 23.

```

=== Detailed Accuracy By Class ===
TP Rate   FP Rate   Precision  Recall   F-Measure  Class
1         0         1          1        1          muyalto
1         0         1          1        1          alto
1         0         1          1        1          medio
1         0         1          1        1          bajo

```

Figura A.15. *Precisión detallada por clase*

El segundo informe *Detailed Accuracy by Class* o precisión detallada por clase muestra los siguientes resultados para cada uno de los valores que puede tomar el atributo de clase: *TP Rate*, *FP Rate*, *Precision*, *Recall*, *F-Measure* y *Class*.

TP Rate es la abreviatura de *True Positives Rate* o clasificación verdadera positiva representa las instancias que son correctamente clasificadas para cada uno de los valores que puede tomar el atributo de clase.

FP Rate es la abreviatura de *False Positives Rate* o clasificación falsa positiva representa las instancias que son incorrectamente clasificadas para cada uno de los valores que puede tomar el atributo de clase.

Las otras columnas *precision*, *recall*, *F-measure* se relacionan con las dos anteriores. Los resultados del informe se observan en la figura A.15.

El tercer informe es Confusion Matrix o matriz de confusión: Es una herramienta de visualización que se emplea en aprendizaje supervisado. Cada columna de la matriz representa el número de predicciones de cada clase, mientras que cada fila representa a las instancias en la clase real. Uno de los beneficios de las matrices de confusión es que facilitan ver si el sistema está confundiendo dos clases. Los valores de la diagonal son los aciertos, y el resto de los valores son los errores, se puede observar en la figura A.16.

```

=== Confusion Matrix ===
a b c d  <-- classified as
4 0 0 0 | a = muyalto
0 5 0 0 | b = alto
0 0 5 0 | c = medio
0 0 0 9 | d = bajo

```

Figura A.16. *Matriz de confusión*

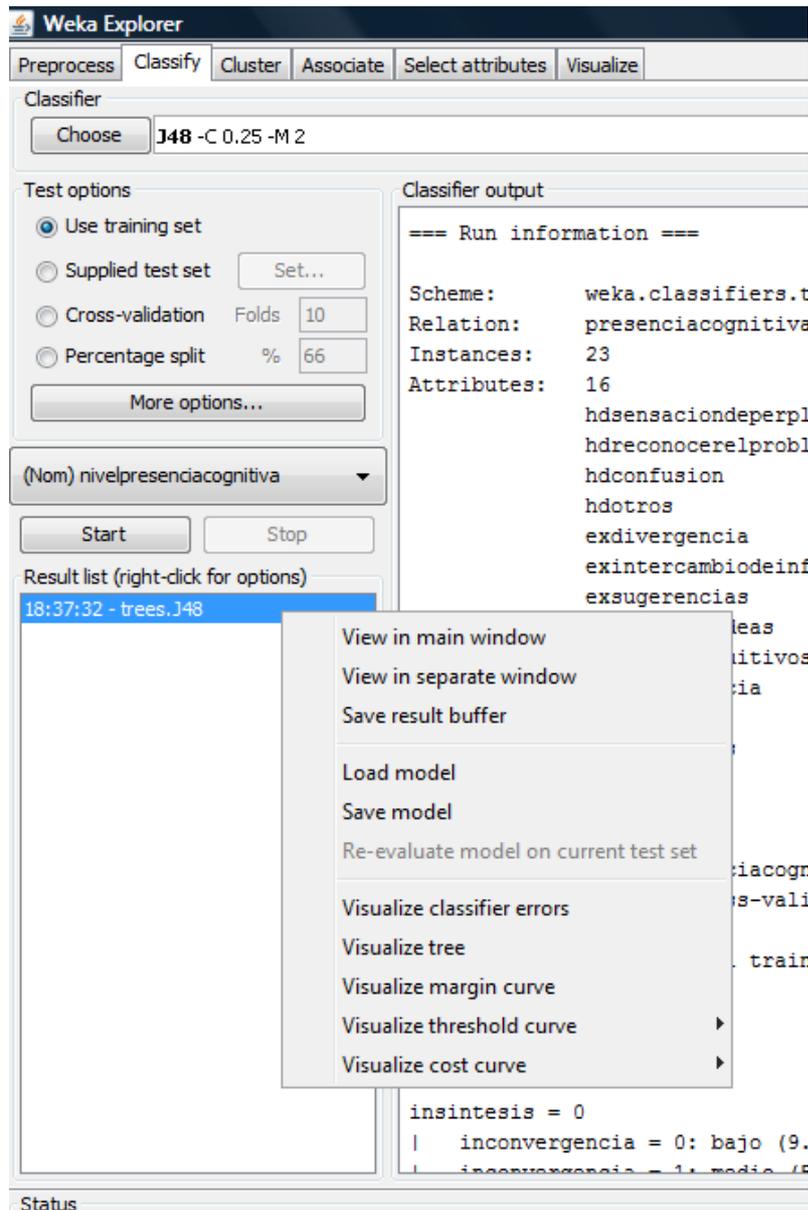


Figura A.17. Lista de resultados

Al igual que con otras opciones de análisis, la ventana izquierda de la lista de resultados contiene el resumen de todas las aplicaciones de clasificadores sobre conjuntos de datos en la sesión del *Explorer*. Puede accederse a esta lista para presentar los resultados, y al activar el botón derecho aparecen diferentes opciones de visualización, entre las que se destacan: *Load model*, *Save model*, *Visualize tree* y *Visualize classifier errors* que se observan en la figura A.17.

Load model, Save model, es decir salvar y cargar modelos. Estos modelos pueden recuperarse de fichero para posteriormente aplicarlos a nuevos conjuntos de datos

Visualize tree o visualizar árbol se puede ver el árbol que generó el clasificador como el que se muestra en la figura A.18. En cada nodo del árbol, el algoritmo J48 elige el atributo de los datos que más eficazmente divide el conjunto de instancias en subconjuntos. Su criterio es el normalizado para ganancia de información (diferencia de entropía) que resulta en la elección de un atributo para dividir los datos. El atributo con la mayor ganancia de información normalizada se elige como parámetro de decisión como se explicó en el Capítulo 3.

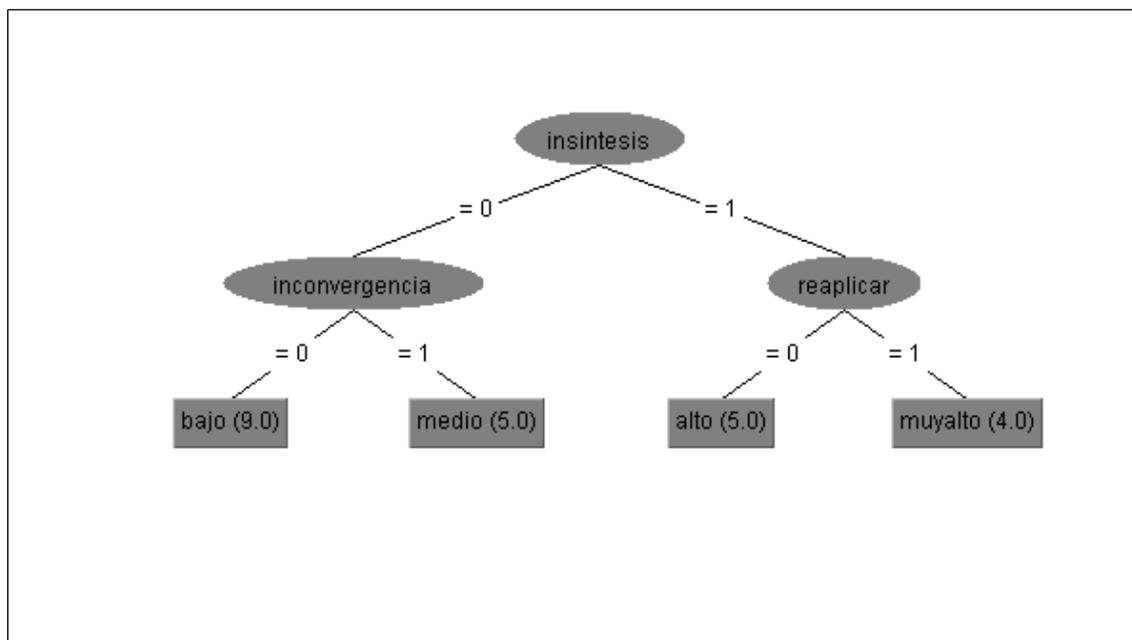


Figura A.18. *Árbol de clasificación*

A.5.3. Búsqueda de asociaciones

La cuarta pestaña *Associate* muestra una ventana que permite aplicar métodos orientados a buscar asociaciones entre datos. Es importante reseñar que estos métodos sólo funcionan con datos enumerados. Éste es sin duda el apartado más sencillo y más simple de manejar, ya que ofrece opciones. Basta con seleccionar un método, por medio del botón *Choose*, configurarlo y verlo funcionar. Entre los algoritmos que ofrece, se encuentra Apriori. En la figura A.19 se puede observar la ventana principal de Associate.

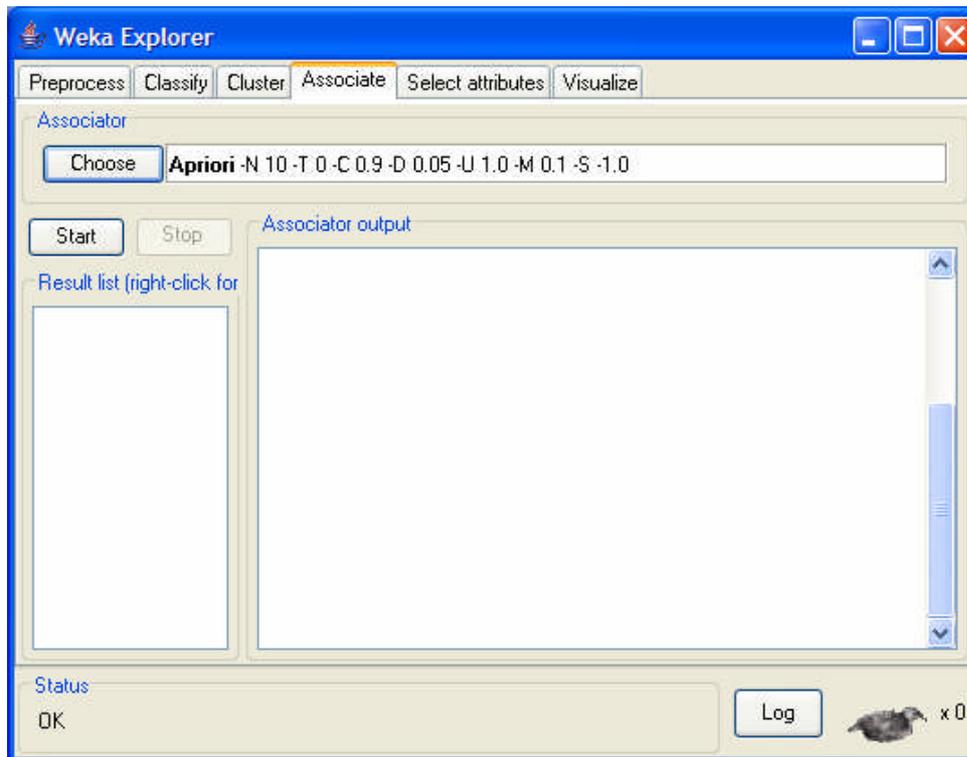


Figura A.19. Ventana principal de Associate

La figura A.20 muestra los parámetros con los que el algoritmo Apriori, selecciona reglas de asociación cuando es ejecutado.

delta: reduce el valor de soporte iterativamente por este factor, hasta que el mínimo soporte o la cantidad de reglas requeridas han sido generadas.

lowerBoundMinSupport: límite inferior para mínimo soporte

metricType: tipo de la métrica para clasificar las reglas: Confidence, Lift, Leverage, Conviction.

minMetric o puntaje métrico mínimo: considera solamente reglas con puntajes más altos que este valor.

numrules: número de reglas a encontrar

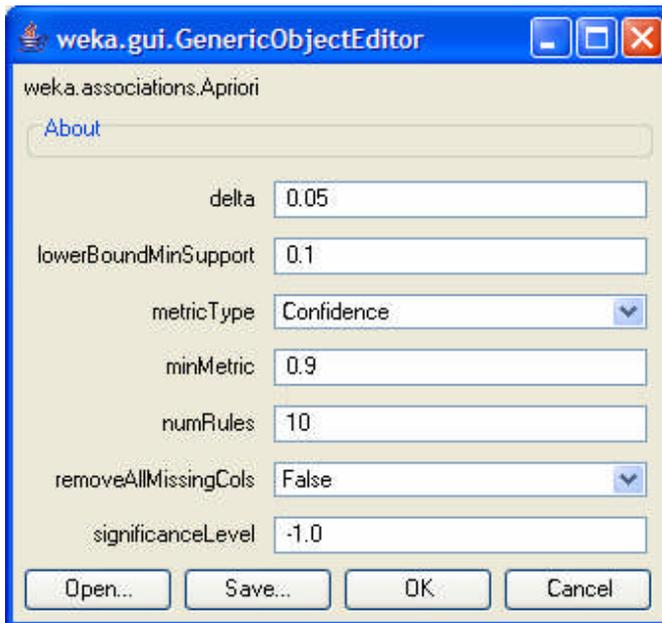


Figura A.20. *Parámetros del algoritmo Apriori*

La figura A.21. muestra las 10 reglas generadas para el archivo presenciacognitiva.arff, para los valores por defecto.

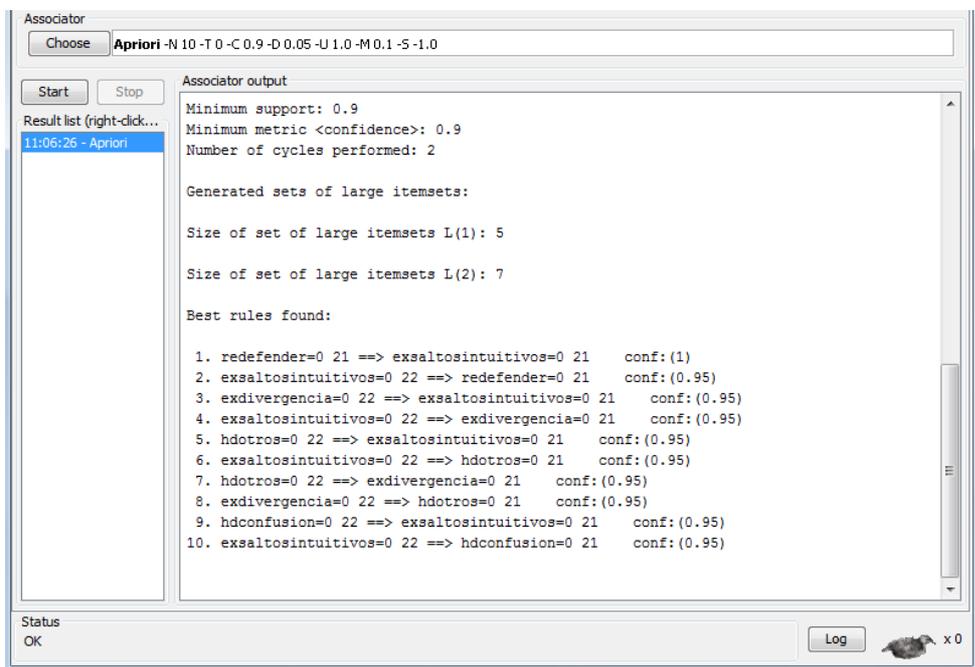


Figura A.21. *Reglas generadas con el algoritmo A priori*

ANEXO B

Instrumentos y Heurística

ANEXO B: INSTRUMENTOS Y HEURÍSTICA

B.1. DEFINICIÓN DE LA UNIDAD DE ANÁLISIS

La unidad de análisis utilizada es el *mensaje completo* y se considera a éste como *el comprobante de la interacción*.

Cada mensaje se divide en partes que pueden ser identificadas en una de las subcategorías de análisis establecido en la tabla B.1.

Un fragmento es una contribución interpretable y codificable en una de las subcategorías del sistema de análisis establecido.

El mensaje completo puede estar formado por uno o varios fragmentos.

B.2. PROCEDIMIENTO DE CODIFICACION

- Se identifican los fragmentos de cada mensaje para una de las dimensiones (social, docente o cognitiva).
- Un mismo fragmento sólo puede ser codificado una vez en una subcategoría de cada dimensión.
- Los distintos fragmentos de un mensaje pueden ser codificados, si procede, en las distintas subcategorías de una misma dimensión.
- A efectos del análisis de reglas de asociación, cuando en un mensaje hay varios fragmentos codificados en la misma categoría de una dimensión, se contabilizan una sola vez.
- A efectos del análisis con árboles de decisión, cuando en un mensaje hay varios fragmentos codificados en la misma categoría de una dimensión, se contabilizan una sola vez si corresponden al mismo objetivo de aprendizaje.

B.3. PREPARACIÓN DE LOS DATOS

- Compilar en un archivo todos los mensajes del foro de cada grupo.(ARCHIVO1)
- Ocultar la identidad de los participantes, tanto tutores como estudiantes, asignando un código identificador.
- Generar otro archivo que relacione el código identificador utilizado con el nombre real del participante, el grupo, el curso, el foro, etc. al que pertenece.(ARCHIVO2)
- Asignar número a cada uno de los mensajes del ARCHIVO1, en el caso que el primer mensaje del foro sea la consigna del trabajo se excluye de la numeración.
- Si el mensaje tiene documentos adjuntos relacionados con los objetivos de aprendizaje del foro, se inserta el contenido textual y completo del documento al mensaje en cuestión, colocando una marca que indique donde comienza el documento y otra marca donde finaliza.
- Se entregan copias del ARCHIVO1 obtenido a cada uno de los docentes que van a codificar.

En el Anexo C se muestra el archivo preparado para la codificación de los mensajes del foro de la práctica de aprendizaje grupal de la unidad I, del grupo G5 del curso de postgrado Prácticas educativas en Entornos Virtuales.

B.4. INSTRUMENTO PARA CATEGORIZAR

El instrumento que se presenta en la tabla B.1. tiene como objetivos contribuir a la precisión en la tarea de categorización de los fragmentos en la *Dimensión Cognitiva* y aumentar el nivel de acuerdo en las revisiones de los codificadores.

Los indicadores de cada subcategoría están definidos con detalle y además, tienen un alto grado de exclusividad y exhaustividad entre los mismos.

Tabla B.1. Indicadores de Presencia Cognitiva

Categoría	Subcategoría	Indicador	Identificador
HECHO DESENCADENANTE	Sensación de perplejidad: cuando existen deficiencias en la adquisición, elaboración y	El estudiante hace referencia a sus conocimientos previos o a	Ha

<p>EVOCATIVO (inductivo)</p> <p><i>Advertir el problema</i>, esto es, el conflicto entre la experiencia pasada (lo sentido, imaginado, visto, comprendido o realizado) de quien aprende y los nuevos hechos, datos, fenómenos o situaciones que se presentan en contradicción con esa experiencia. Lo nuevo que se aprende del mundo, se aprende relacionándolo (reflexionando) con la experiencia anterior. Se trata de conceptualizar el problema, por esta razón se lo considera como evocativo e inductivo por naturaleza.</p> <p>En esta primera fase, el proceso educativo se basa principalmente en la presentación de información que genere curiosidad y preguntas, para que la indagación avance progresivamente.</p>	<p>expresión de lo conocido o en las habilidades o competencias del sujeto que aprende. Implica un estado de duda, de perplejidad, de dificultad unido a una voluntad que desea encontrar una solución a un problema.</p> <p>Surge una dificultad o problema que hace volver a la mente (reflectio) sobre el tema u objeto ya conocido.</p> <p>La razón suficiente y necesaria para justificar la reflexión es el deseo de buscar una solución a una perplejidad o problema.</p>	<p>sus dudas sobre la temática en relación con la consigna.</p>	
	<p>Reconocer el problema: se vuelve sobre lo conocido, se reflexiona, porque lo ya conocido no es lógica y psicológicamente suficiente para explicar lo nuevo, para poder actuar o producir algo en base a lo ya conocido. Se localiza y delimita la dificultad por comparación o confrontación de aspectos, objetos, situaciones, etc.</p>	<p>El estudiante localiza y delimita el problema y expone sus ideas fundamentadas en la bibliografía propuesta para la consigna.</p>	Hb
	<p>Confusión: no se localiza o no se delimita el problema.</p>	<p>El estudiante hace referencia a temas que no están relacionados con la consigna expresándolos como pertinentes al problema planteado.</p>	Hc
	<p>Otros (Contingencia) es un concepto que procede del latín <i>contingentia</i>. El término suele referirse a algo que es probable que ocurra, aunque no se tiene una certeza al respecto. La contingencia, por lo tanto, es lo posible o aquello que puede, o no, concretarse. Situaciones accidentales que impiden transitoriamente las actividades cognitivas.</p>	<p>El estudiante hace referencia a contingencias que no están relacionados con la consigna expresándolos como accesorios al problema planteado. Por ejemplo problemas relacionados con temas técnicos o con las tecnologías (no pude ver el video, no pude acceder a la plataforma, tengo un servicio de internet muy</p>	Hd

		lento, etc.).	
<p>EXPLORACIÓN</p> <p>INQUISITIVO (<i>divergente</i>)</p> <p>Implica entender la naturaleza del problema y buscar información relevante y explicaciones posibles.</p> <p>Las ideas son exploradas de modo cooperativo y los individuos intentan encontrar sentido a la aparente complejidad y confusión. Se trata, por tanto, de un proceso inquisitivo y divergente. El proceso educativo, en esta segunda fase, implica lanzar y presentar ideas para corroborar o para oponerse a otras, recibir nuevas perspectivas, explicar experiencias y hacer observaciones y comentarios sobre el valor de toda esa información.</p>	<p>Divergencia: se discrepa, se separa, se plantean alternativas o puntos de vista diferentes para la resolución de un problema.</p>	El estudiante hace comentarios fundamentados que difieren o presentan otro punto de vista respecto a las ideas expresadas en la bibliografía propuesta, o con las expuestas en las intervenciones de otros participantes en relación con la consigna planteada.	Ea
	<p>Intercambio de información: se comparte información, documentación experiencias, etc. con los demás participantes.</p>	El estudiante aporta documentos, direcciones electrónicas, distinto tipo de información o comparte experiencias que agregan conocimiento sobre la temática en relación con la consigna.	Eb
	<p>Sugerencias: se proponen ideas, se realizan observaciones o comentarios que contribuyen a la resolución de un problema.</p>	El estudiante propone ideas, realiza observaciones o comentarios fundamentados acerca del valor de un aporte que agrega conocimiento sobre la temática en relación con la consigna.	Ec
	<p>Lluvia de ideas: es una técnica de trabajo grupal que facilita el surgimiento de nuevas ideas sobre un tema o problema determinado, para generar ideas originales en un ambiente relajado. En un principio toda idea es válida y ninguna debe ser rechazada.</p>	El estudiante propone nuevas ideas con el objetivo que se hagan observaciones sobre el valor de las mismas en relación con la consigna.	Ed
	<p>Saltos intuitivos: en contraste con el pensamiento analítico, el pensamiento intuitivo no avanza por pasos cuidadosos y bien definidos. En realidad, tiende a incluir</p>	El estudiante propone ideas, realiza comentarios, observaciones en forma intuitiva omitiendo la fundamentación de los	Ee

	<p>maniobras basadas aparentemente en una percepción implícita de la totalidad del problema. El pensador llega a una respuesta que, puede ser correcta o incorrecta, con muy poca o ninguna conciencia del proceso mediante el cual llegó a ella. Muy rara vez puede proporcionar una relación adecuada de la manera en que obtuvo su respuesta, y suele no darse cuenta de, exactamente, a cuáles aspectos de la situación problemática estaba respondiendo. Por lo general, el pensamiento intuitivo se basa en una familiaridad con el dominio del conocimiento implicado, y con su estructura, lo que hace posible para el pensador “saltar”, omitiendo pasos y empleando “atajos” en forma que requerirá una comprobación ulterior de conclusiones por medios más analíticos, ya sean deductivos o inductivos.</p>	temas en relación con la consigna.	
<p>INTEGRACIÓN</p> <p>TENTATIVO (<i>convergente</i>)</p> <p>Se orienta hacia la construcción del significado. Es el proceso de elaborar una solución o explicación apropiada. Por eso se considera como una asociación provisional de ideas (tentativo) capaz de ofrecer significado y de ofrecer soluciones potenciales. El proceso educativo en la tercera fase incluiría,</p>	<p>Convergencia Se toman decisiones sobre la integración de ideas y sobre la sistematización progresiva de las mismas. Se trata de integrar información, intercambiar mensajes que manifiesten acuerdo, construir sobre la base de otras ideas, presentar una explicación, y ofrecer justificación</p>	El estudiante analiza las ideas propuestas, manifiesta acuerdos con otras contribuciones y va organizando la información progresivamente.	Ia
	<p>Síntesis: es el proceso mediante el cual combinamos elementos diversos para formar un todo coherente. El análisis y la síntesis son dos procesos complementarios, el primero va del todo a las partes</p>	El estudiante combina las ideas propuestas para formar un todo coherente tratando de construir una explicación al problema planteado.	Ib

<p>por tanto, los siguientes aspectos: integrar información, intercambiar mensajes que manifiesten acuerdo, construir sobre la base de otras ideas, presentar una explicación, y ofrecer explícitamente una solución.</p> <p>Aunque se trata de una fase básicamente reflexiva, los estudiantes deben implicarse activamente en el discurso crítico que debe configurar la comprensión.</p>	<p>y el segundo de las partes al todo. Si bien desde el punto de vista exclusivamente lógico puede considerarse al análisis como previo al proceso de síntesis, en la práctica ambos procesos están íntimamente entrelazados.</p>		
	<p>Soluciones: surgimiento creativo de una posible solución. La reflexión se vuelve sobre un aspecto, objeto o situación posible en la que el problema presente carece de sentido y se plantea una solución teórica (de interpretación), práctica (de acción), o productiva (generar productos, hechos). Explicar el problema: sacar las conclusiones que se derivan, con coherencia, de las teorías o hipótesis elaboradas</p>	<p>El estudiante construye en base de las ideas analizadas y sintetizadas una solución potencial al problema planteado.</p>	<p>Ic</p>
<p>RESOLUCIÓN</p> <p>COMPROMETIDO <i>(deductivo)</i></p> <p>Es la resolución del dilema o problema, sea reduciendo la complejidad mediante un marco de significado, sea descubriendo una solución contextualmente específica.</p> <p>Se debe evaluar la viabilidad de la solución propuesta mediante su aplicación directa o indirecta.</p> <p>El proceso educativo, en la cuarta fase, requiere la prueba deductiva de la solución, que puede realizarse mediante una implementación directa o</p>	<p>Aplicar se prueba la solución propuesta para evaluar su viabilidad.</p> <p>La prueba podría consistir en una aplicación directa o en un proyecto de investigación-acción (individual o colectiva). En el caso de la aplicación indirecta o experimental se requiere un análisis riguroso de la prueba hipotética, que podría consistir en una presentación expuesta al juicio de los demás participantes.</p>	<p>El estudiante aplica la solución propuesta y la expone al juicio de los demás participantes.</p>	<p>Ra</p>
	<p>Comprobar: verificar las consecuencias posibles derivadas de la solución aplicada. Confrontar los resultados obtenidos con los buscados.</p>	<p>El estudiante comprueba los beneficios y consecuencias de la aplicación de la solución propuesta y expone los resultados al juicio de los demás participantes.</p>	<p>Rb</p>
	<p>Defender: En el caso de la aplicación indirecta o experimental se requiere un</p>	<p>El estudiante defiende la solución propuesta y expone su interpretación</p>	<p>Rc</p>

<p>indirecta. Por otra parte, en los buenos entornos docentes, al igual que en la vida real, es muy poco frecuente dar con soluciones definitivas. Inevitablemente, los resultados de la fase de resolución suelen plantear nuevas preguntas, activando nuevos ciclos de indagación y, por tanto, promoviendo el aprendizaje continuo.</p>	<p>análisis riguroso de la prueba hipotética, que podría consistir en una defensa expuesta al juicio de los demás participantes.</p>	<p>de los resultados al juicio de los demás participantes.</p>	
--	--	--	--

B.5. PROCEDIMIENTO PARA CATEGORIZAR

Tres analistas, de manera independiente codifican los datos preparados en el ARCHIVO1 y los vuelcan en el instrumento mostrado en la tabla B.2.

La categorización de los mensajes se realiza en base al procedimiento de codificación (inciso B.2.) y aplicando las definiciones de la tabla B.1. para los Indicadores de Presencia Cognitiva.

La revisión de la categorización tiene lugar de manera compartida entre los investigadores que discuten sus dificultades para asignar cada fragmento a alguna categoría

Durante las sesiones convocadas se utiliza un control de fiabilidad por consenso entre los tres investigadores, donde los registros se comparan para controlar el nivel de acuerdo existente.

En base a los acuerdos se obtienen los resultados que se vuelcan en el instrumento descrito en la tabla B.2.

B.6. INSTRUMENTOS PARA RECOLECCIÓN

B.6.1. Reglas de asociación

Los datos de la categorización a analizar con las reglas de asociación se estructuran en una tabla. Las *filas* de la tabla se refieren a los *mensajes* y las *columnas* a cada uno de

indicadores de presencia cognitiva definidos. La tabla solo contiene valores binarios. Un 1 en la posición (i,j) indica que el mensaje i contiene el indicador j, mientras que un 0 indica que el mensaje no ha mostrado el indicador.

El instrumento contiene además de las columnas de los indicadores, tres columnas que corresponden a emisor, grupo e identificador del mensaje, respectivamente. Se muestra en la tabla B.2. la estructura del instrumento y los datos corresponden a la categorización de los mensajes del foro de la práctica de aprendizaje grupal de la unidad I, del grupo G5 del curso de postgrado “Prácticas educativas en Entornos Virtuales” revisado por tres analistas.

Tabla B.2. *Instrumento para recolección de indicadores por mensaje*

Emisor	Identificador del Mensaje	Ha	Hb	Hc	Hd	Ea	Eb	Ec	Ed	Ee	Ia	Ib	Ic	Ra	Rb	Rc
E1	M1	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0
E2	M3	0	1	0	0	0	1	1	0	0	0	1	0	0	0	0
E3	M4	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
E4	M5	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
E4	M8	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0
E1	M10	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
E3	M15	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
E1	M16	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
E2	M19	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0
E4	M20	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
E4	M23	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0
E2	M24	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
E1	M25	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
E3	M28	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
E4	M29	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
E3	M30	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
E1	M33	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
E4	M34	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
TOTALES		18	4				6	1			11	6				
CONTROL			4				6	1			11	6				

En la tabla B.2. están los datos preprocesados, para utilizarlos con herramientas automatizadas habrá que transformar el formato de acuerdo a los requerimientos del software. (Ver Anexo A).

B.6.2. Árboles de Decisión

En forma análoga a la aplicada para las reglas de asociación, los valores de los indicadores de presencia cognitiva observados en los foros de aprendizaje se organizan en una tabla. En este caso, cada fila representa *un estudiante* que participa en el foros, las columnas son los indicadores de presencia cognitiva a alcanzar *para un objetivo de aprendizaje*. La tabla sólo contiene valores binarios. Un 1 en la posición (i,j) indica que el estudiante i ha alcanzado el indicador j, mientras que un 0 indica que el estudiante no ha logrado el indicador.

Se generan las tablas, una por cada estudiante, con los mensajes correspondientes al objetivo de aprendizaje que se evalúa. Las nuevas tablas son similares a las usadas para reglas de asociación y se les agregan dos filas una correspondiente a los totales y la otra con el resumen.

La fila de totales sirve para generar la fila resumen y se utiliza como control una vez que se tienen las tablas de todos los estudiantes. La fila resumen contiene valores binarios, 0 si el estudiante no logro el indicador en ninguno de sus mensajes, y 1 si alcanzó el indicador en al menos 1 mensaje, es decir el haber logrado el indicador se contabiliza una sola vez para el objetivo de aprendizaje que se evalúa.. (Ver tablas B.3., B.4., B.5. y B.6.)

Tabla B.3. Instrumento para recolección de datos estudiante E1

Emisor	Identificador del Mensaje	Ha	Hb	Hc	Hd	Ea	Eb	Ec	Ed	Ee	Ia	Ib	Ic	Ra	Rb	Rc
E1	M1	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0
E1	M10	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
E1	M16	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
E1	M25	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
E1	M33	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
E1	Total	0	1	0	0	0	1	0	0	0	3	2	0	0	0	0
E1	Resu	0	1	0	0	0	1	0	0	0	1	1	0	0	0	0

Tabla B.4. Instrumento para recolección de datos estudiante E2

Emisor	Identificador del Mensaje	Ha	Hb	Hc	Hd	Ea	Eb	Ec	Ed	Ee	Ia	Ib	Ic	Ra	Rb	Rc
E2	M3	0	1	0	0	0	1	1	0	0	0	1	0	0	0	0

E2	M19	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0
E2	M24	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
E2	Total	0	1	0	0	0	1	1	0	0	2	2	0	0	0	0
E2	Resu	0	1	0	0	0	1	1	0	0	1	1	0	0	0	0

Tabla B.5. Instrumento para recolección de datos estudiante E3

Emisor	Identificador del Mensaje	Ha	Hb	Hc	Hd	Ea	Eb	Ec	Ed	Ee	Ia	Ib	Ic	Ra	Rb	Rc
E3	M4	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
E3	M15	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
E3	M28	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
E3	M30	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
E3	Total	0	2	0	0	0	2	0	0	0	2	0	0	0	0	0
E3	Resu	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0

Tabla B.6. Instrumento para recolección de datos estudiante E4

Emisor	Identificador del Mensaje	Ha	Hb	Hc	Hd	Ea	Eb	Ec	Ed	Ee	Ia	Ib	Ic	Ra	Rb	Rc
E4	M5	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
E4	M8	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0
E4	M20	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
E4	M23	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0
E4	M29	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
E4	M34	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
E4	Total	0	0	0	0	0	2	0	0	0	4	2	0	0	0	0
E4	Resu	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0

Para completar la tabla resta definir un resultado o “función objetivo” que se obtiene del conjunto de valores observados, de acuerdo con un dominio de conocimiento que se quiere adquirir. Este resultado para el dominio que estamos analizando es el “Nivel de Presencia Cognitiva” (NPC) alcanzado por el estudiante para un objetivo de aprendizaje, y tendrá los valores: Muy Alto, Alto, Medio o Bajo.

Para evaluar el nivel de presencia cognitiva alcanzado por cada estudiante se sigue un procedimiento similar al de categorización.

Tres tutores, de manera independiente definen el nivel de presencia cognitiva de cada estudiante para un objetivo de aprendizaje.

La revisión de la evaluación tiene lugar de manera compartida entre los tutores.

Durante las sesiones de revisión se utiliza un control de fiabilidad por consenso entre los tres tutores, donde los registros se comparan para controlar el nivel de acuerdo existente.

En base a los acuerdos se obtienen los resultados se vuelcan en un instrumento similar a los anteriores al que se agrega la columna con la evaluación, como el de la tabla B.7.

Tabla B.7. Instrumento para evaluación nivel de presencia cognitiva

Emisor	Identificador del Mensaje	Ha	Hb	Hc	Hd	Ea	Eb	Ec	Ed	Ee	Ia	Ib	Ic	Ra	Rb	Rc	NPC
E1	Resu	0	1	0	0	0	1	0	0	0	1	1	0	0	0	0	Alto
E2	Resu	0	1	0	0	0	1	1	0	0	1	1	0	0	0	0	Alto
E3	Resu	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0	Medio
E4	Resu	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	Alto

En la tabla B.7 están los datos preprocesados, para utilizarlos con herramientas automatizadas habrá que transformar el formato de acuerdo a los requerimientos del software. (Ver Anexo A).

ANEXO C

Foro Grupo G5

