I Escuela de Verano en Inteligencia Artificial
Coruña, Septiembre de 2014

Introducción a la Minería de Datos Educativos

Sebastián Ventura

Department of Computer Sciences and Numerical Analysis. University pf Córdoba
Department of Computer Sciences. King Abdulaziz University

Outline

- Introducción
- Procesos y actores implicados en EDM
- Objetivos en EDM
- Tipos de datos educativos
- Tareas en EDM
  - Tareas de bajo nivel
  - Tareas de alto nivel
- Algunos casos de éxito
- Nuevos retos / Problemas abiertos
- Recursos interesantes

Introducción

Introducción

- El desarrollo de sistemas de enseñanza basada en web se ha incrementado exponencialmente en los últimos años.
  - Estos sistemas generan información de gran valor pedagógico, pero suele ser tan abundante que resulta imposible analizarla manualmente.
  - Se necesitan herramientas que sean capaces de analizar esos datos de forma **automática**.
- Las instituciones educativas disponen de sistemas de información con gran cantidad de datos interesantes.
  - La información disponible en estos sistemas puede utilizarse para mejorar el plan estratégico de la institución.
  - En este caso, también se necesitan herramientas que analicen esos datos de forma **automática.**

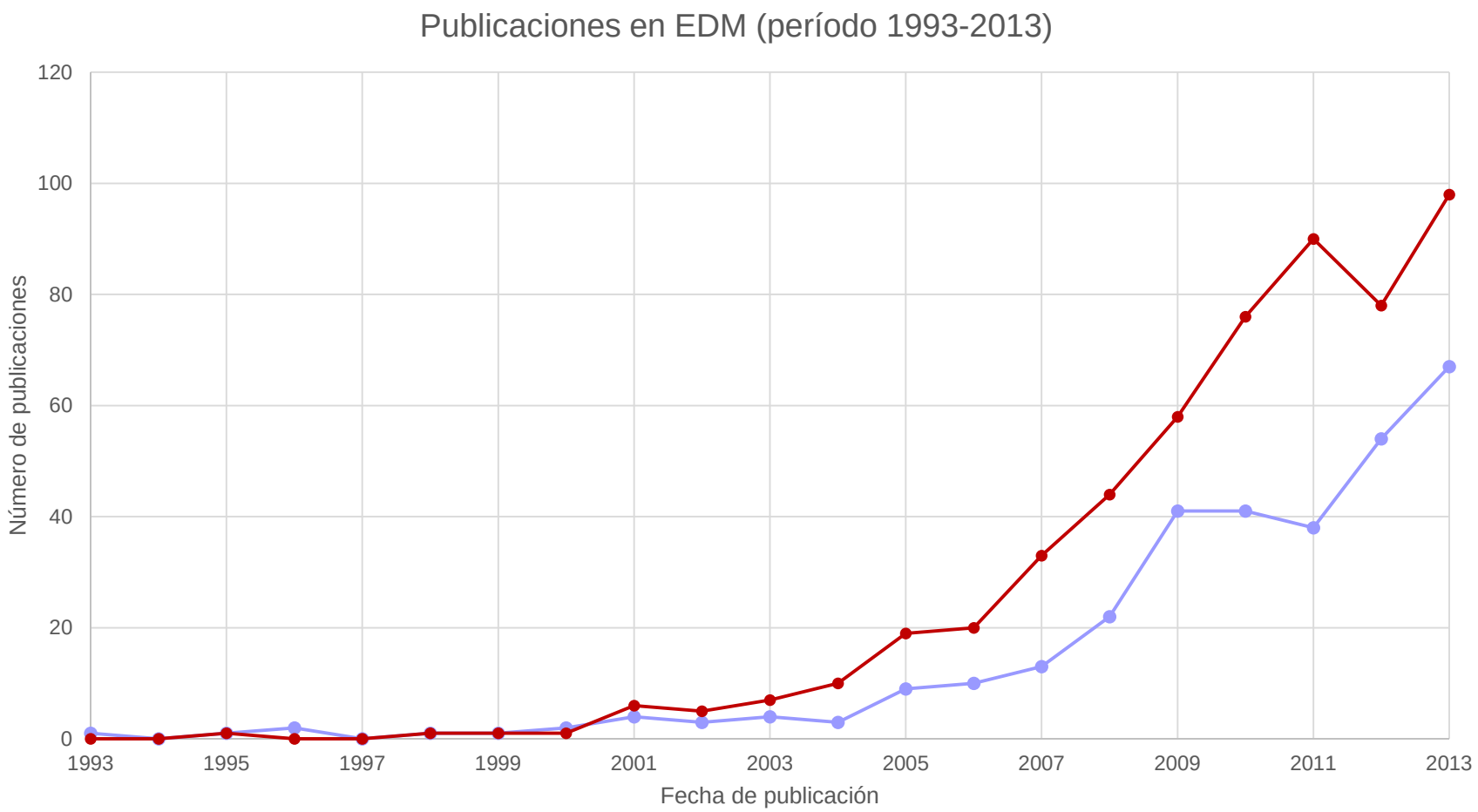
Introducción

- Se denomina minería de datos educativos (***educational data mining, EDM)*** a la aplicación de técnicas de minería de datos a información generada en los entornos educativos.
- Las primeras referencias del área datan del 1995.
- Se ha experimentado un crecimiento notable en las publicaciones sobre el tema en los últimos años.
- También existe un grupo de trabajo internacional sobre investigación en minería de datos educativos

  http://www.educationaldatamining.org/index.html
- Desde 2008 se celebra anualmente la *International Conference on Educational Data Mining*, que representa el referente dentro de esta línea de trabajo.

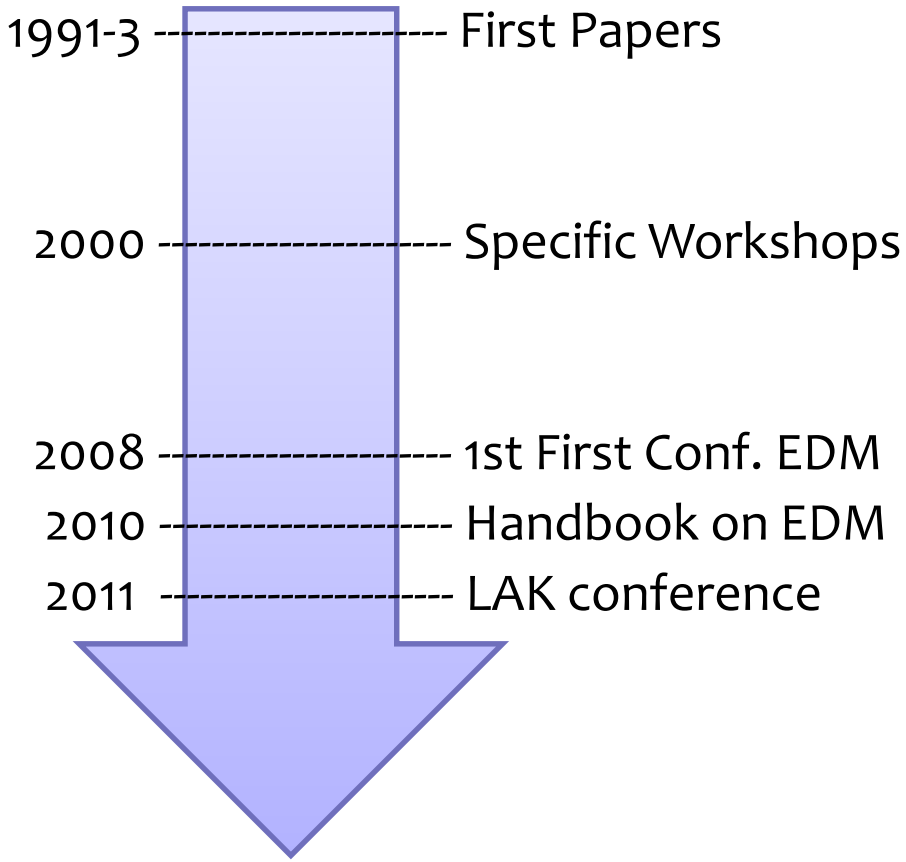
Introducción

Publicaciones EDM (SCOPUS)



Introducción

Evolución histórica



Introducción

Dos áreas íntimamente relacionadas con EDM

- **Analítica del aprendizaje (Learning Analytics)**

  Medida, colección, análisis y generación de informes sobre estudiantes y sus contextos, con la intención de comprender y optimizar el aprendizaje y el entorno en el que tiene lugar.

- **Analítica académica (Academic Analytics)**

  Aplicación de las técnicas de inteligencia de negocio (*Business Intelligence*) a datos académicos institucionales.

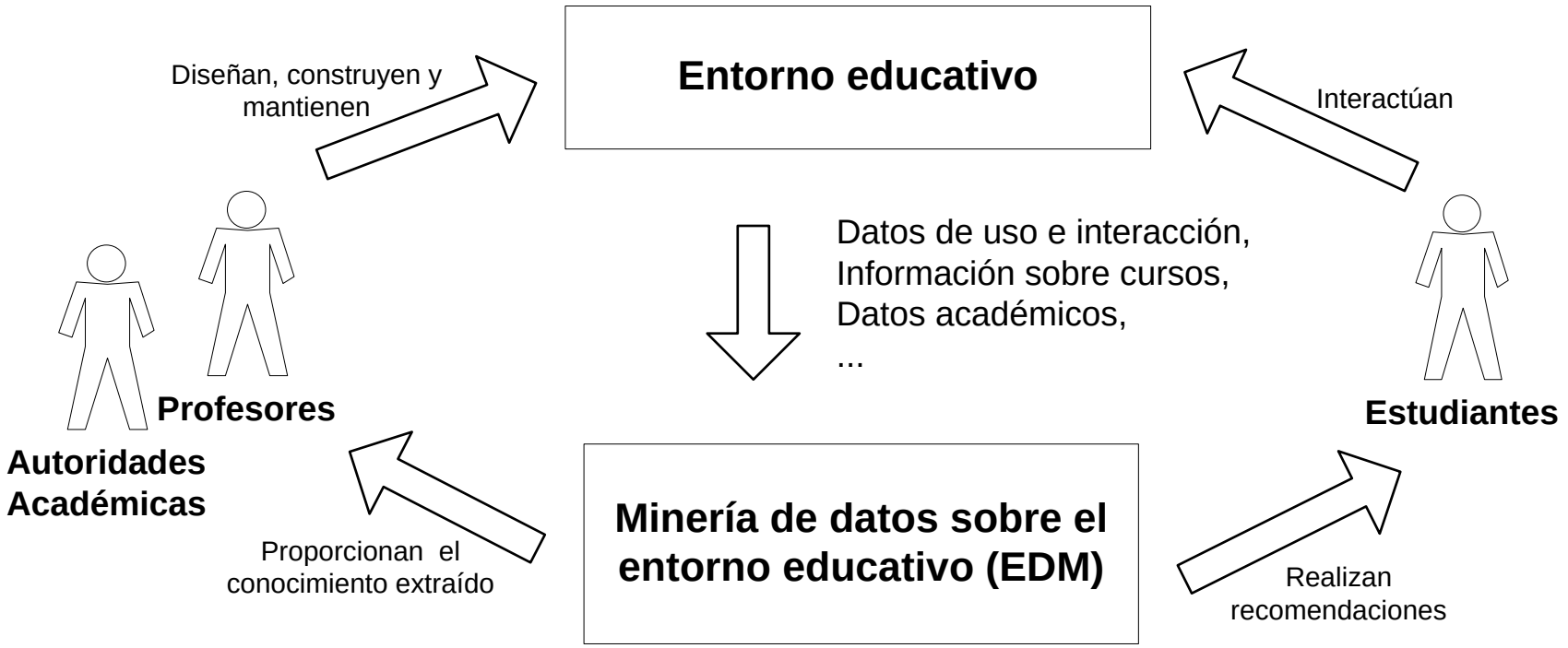
Introducción

Dos áreas íntimamente relacionadas con EDM (II)

- Realmente, la diferencia entre EDM, LA y AA se debe más a los orígenes de los equipos que trabajan en estos temas que a otra cuestión
- EDM, al igual que LA, tiene como uno de sus objetivos comprender y optimizar el aprendizaje y el entorno en el que tiene lugar.
- Algunas de las primeras referencias de EDM hacían uso de datos académicos, y planteaban decisiones estratégicas como en AA.
- Lo más correcto sería definir una única disciplina denominada **Educational Data Science (EDS)**.

Proceso y actores en EDM

EDM: Proceso y actores



Objetivos de EDM

Objetivos del EDM

- El conocimiento que puede extraerse de los sistemas educativos es muy diverso.
- El objetivo que nos marcamos al intentar aplicar técnicas de EDM depende de:
  - **A quién** va dirigido el conocimiento que extraigamos
    - Alumnos
    - Profesores
    - Autoridades académicas
  - De qué **tipo de información** disponemos
    - A priori
    - A posteriori
  - **Entorno** en el que nos situamos
    - Enseñanza presencial
    - Enseñanza a distancia
  - ...

Objetivos del EDM

Interés del estudiante

- Qué actividades, recursos y tareas podrían mejorar el rendimiento académico de los alumnos.
- Qué actividades se ajustan mejor al perfil de un determinado alumno.
- Qué camino recorrer para obtener un resultado concreto:
  - Basándonos en conocimiento del camino ya recorrido por el alumno y su éxito.
  - Por comparación con lo realizado por otros alumnos de características análogas.

Objetivos del EDM

Interés del profesor

- Cuantificar la efectividad del proceso de enseñanza-aprendizaje
- Organizar los contenidos de un curso
- Mejorar o corregir la estructura del curso
- Clasificar o agrupar alumnos en base a sus características
  - Tutorización y asesoramiento
  - De cara a monitorizar conocimiento interesante
- Buscar patrones de comportamiento en alumnos
  - Patrones generales
  - Patrones anómalos
- Evaluar las actividades realizadas en un curso
  - Efectividad
  - Motivación
- Monitorizar actividades:
  - Errores más frecuentes en la realización de actividades
  - Grado de dificultad de una actividad
- Personalizar y adaptar el contenido de cursos
  - Diseñar planes de instrucción

Objetivos del EDM

Interés de las instituciones educativas

- Mejora de la eficiencia del sitio web y adaptación de éste a los hábitos de sus usuarios:
  - Tamaño de servidor óptimo
  - Distribución de tráfico en la red
- Organización de los recursos institucionales
  - Diseño de horarios
  - Adquisición de material
- Mejora de la oferta educativa
  - Programas orientados a demanda
  - Orientación de alumnos en base a
    - Objetivos
    - Capacidades

Tareas en EDM

Tareas en EDM

- **Tareas de bajo nivel**. Similares a las tareas en DM convencional, aunque el conocimiento a descubrir se extraerá de los datos educativos.

GENERAL (DM)

- **Tareas de alto nivel**. Intentan resolver un problema en un contexto educativo. Implican la ejecución de una o más tareas de bajo nivel, así como la interpretación y/o validación de resultados.

ESPECÍFICO (EDM)

Tareas en EDM

Tareas de bajo nivel

- **Tareas predictivas**
  - Supervisadas. Se dispone de información de salida
  - Ejemplos
    - Clasificación
    - Regresión
- **Tareas descriptivas**
  - No supervisadas. No se dispone de información de salida
  - Ejemplos
    - Asociación
    - Agrupamiento (clustering)

Tareas en EDM

Tareas de bajo nivel: Clasificación

- Clasificar consiste en identificar a qué conjunto de categorías pertenece una nueva observación a partir de un conjunto de ejemplos en los que las categorías a las que pertenece cada ejemplo son conocidas (método de aprendizaje supervisado).
- Ejemplo: *Construir un modelo para predecir si un alumno aprobará o no una determinada asignatura a partir de una determinada información*.
  - Para llevar a cabo esta tarea…
    - Tenemos información sobre estudiantes que han sido evaluados previamente y que han sido etiquetados como “aprobado” o “suspenso”. Estos ejemplos pueden contener distinto tipo de información.
    - Construimos un modelo usando un algoritmo de clasificación.
    - El modelo nos permitirá predecir si un nuevo estudiante aprobará o suspenderá a partir de su información disponible, proporcionada al modelo.

Tareas en EDM

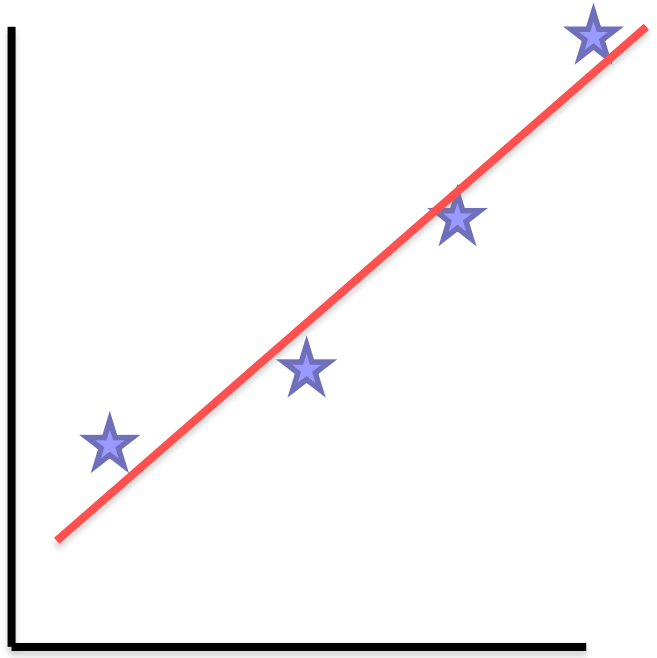
Tareas de bajo nivel: Clasificación (II)

- Predecir el rendimiento de los estudiantes y su nota final (Minaei-Bidgoli & Punch, 2003, Romero et al, 2008)
- Identificar estudiantes que hacen mal uso de las instalaciones (Baker et al., 2004).
- Clasificar estudiantes en (a) guiados mediante consejos y (b) mediante fallos y encontrar los conceptos mal aprendidos con mayor frecuencia (Yudelson et al., 2006).
- Identificar estudiantes con baja motivación y proponer acciones para evitar el abandono académico (Cocca & Weibelzahl, 2006).

Tareas en EDM

Tareas de bajo nivel: Regresión

- El objetivo en regresión es similar al de clasificación, pero en este caso la variable que hemos de aprender es numérica en lugar de categórica.
- Ejemplo: Podemos desarrollar modelos de regresión para predecir las calificaciones numéricas de un conjunto de estudiantes.



Tareas en EDM

Tareas de bajo nivel: Agrupamiento (clustering)

- Consiste en agrupar un conjunto de objetos de manera que los objetos pertenecientes al mismo grupo (denominado **cluster**) sean más parecidos entre sí que los pertenecientes a los demás grupos.
- Ejemplo: Definir grupos de estudiantes similares a partir de la información de uso tomada de un sistema de enseñanza virtual
  - Para realizar esta tarea…
    - Tenemos un conjunto de datos **no etiquetados**.
    - El algoritmo de agrupamiento buscará semejanzas entre los datos y definirá grupos de estudiantes con características similares.
    - El modelo final incluirá la descripción de los grupos resultantes.

Tareas en EDM

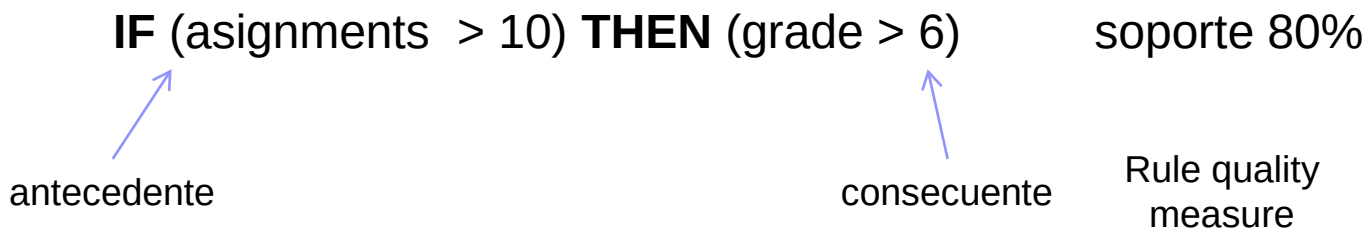
Tareas de bajo nivel: Agrupamiento (II)

- Descubrir patrones que reflejan comportamientos similares en estudiantes, de forma que cuando éstos sean incluidos en espacios de cooperación comunes se obtenga un incremento de su actividad (Talavera & Gaudioso, 2004).
- Agrupar estudiantes para crear itinerarios educativos personalizados para cada grupo (Mor and Minguillon, 2004).
- Agrupar estudiantes según sus capacidades y otras características para personalizar la tutoría (Hamalainen et al., 2004).
- Agrupar estudiantes para fomentar un aprendizaje colaborativo basado en grupos (Tang & McCalla, 2005).
- Agrupar cuestiones y tests en grupos a partir de los datos de una matriz de puntuaciones (Spacco et al., 2006).

Tareas en EDM

Tareas de bajo nivel: Asociación

- Consiste en obtener reglas que asocien conceptos entre diferentes columnas (atributos) de una base de datos.
- El conocimiento extraído tiene forma de reglas, que pueden aplicarse sobre una fracción de los ejemplos almacenados en la base de datos.



- Tarea relacionada: Descubrimiento de subgrupos

Tareas en EDM

Tareas de bajo nivel: Asociación (II)

- Descubrir relaciones interesantes entre la información generada por los estudiantes en un sistema de hipermedia adaptativo y las reglas que lo controlan (Romero et al., 2004).
- Guiar automáticamente la actividad estudiantil y generar y recomendar automáticamente materiales didácticos (Lu, 2004).
- Búsqueda de errores estudiantiles que ocurren con más frecuencia (Merceron & Yacef, 2004).
- Detectar patrones de eventos en equipos estudiantiles, orientados a la detección temprana de errores (Kay et al., 2006).
- Recomendación de hilos de discusión en foros en función de características y/o preferencias de los estudiantes (Abel et al., 2008)

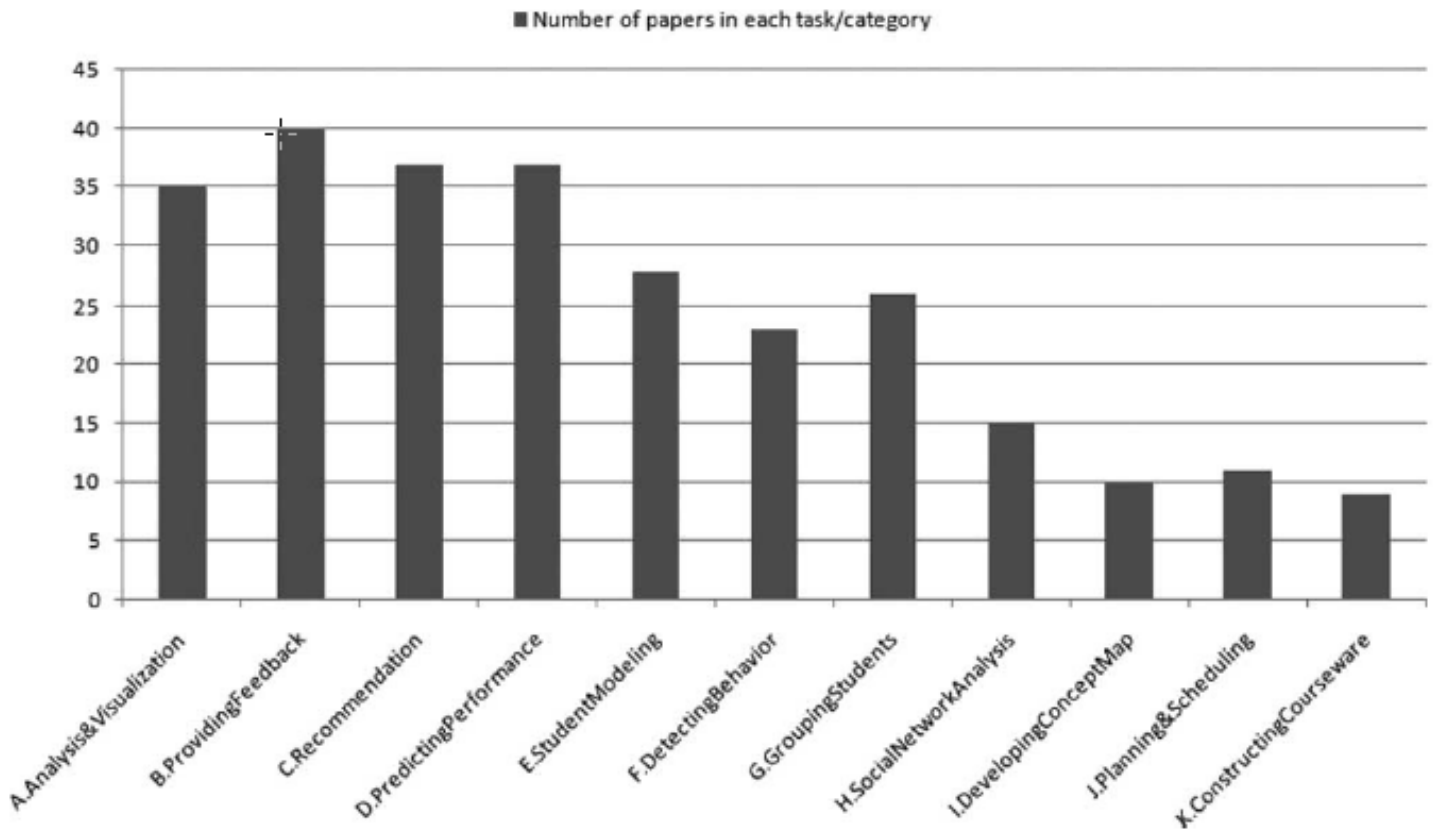
Tareas en EDM

Tareas de alto nivel

- Su objetivo es la resolución de un problema concreto en el ámbito educativo
- Suelen requerir un proceso completo de extracción de conocimiento y, con frecuencia, más de una tarea básica (bajo nivel)

Tareas en EDM

Tareas de alto nivel



Tareas en EDM

Tareas de alto nivel: Predicción del rendimiento académico

- Consiste en estimar el valor de una variable que describe el rendimiento futuro del estudiante a partir de cierta información disponible:
  - Información histórica (evaluaciones previas).
  - Otra información relacionada (social, actitudinal, ambiental…).
- Es una tarea de gran interés, que tiene multiples usos.
  - Llevar a cabo acciones correctivas para mejorar el rendimiento del estudiante, especialmente cuando hay posibilidad de fracaso.
  - Detectar factores críticos para mejorar el rendimiento estudiantil y/o para evitar sus fallos.

Tareas en EDM

Tareas de alto nivel: Predicción del rendimiento académico (II)

- Esta tarea ha sido abordada mediante técnicas de clasificación y regresión:
  - *Clasificación*: Cuando la variable asociada al rendimiento académico es categórica (por ejemplo “aprobado”/”suspenso”)
  - *Regresión*: Cuando la variable es numérica (calificación numérica, número de fallos, etc.)
- Temas abiertos en este campo:
  - Una major evaluación de los modelos de predicción
  - Predicción temprana

Tareas en EDM

Tareas de alto nivel: Recomendar recursos o actividades a los estudiantes

- Consiste en generar nuevo conocimiento que pueda ser utilizado para hacer recomendaciones tales como la siguiente tarea, visita o problema a realizar.
- Este conocimiento puede también ser utilizado para restringir el contenido, los interfaces y las secuencias de aprendizaje de cada alumno individualmente.
- Nos permite personalizar ciertos aspectos del proceso de enseñanza aprendizaje, lo cual es muy conveniente en sistemas de educación a distancia

Tareas en EDM

Tareas de alto nivel: Recomendar recursos o actividades a los estudiantes (II)

Métodos basados en contexto: Analizan la información disponible y construyen un modelo que indica si un determinado recurso es apropiado para un estudiante o grupo de ellos

- *Métodos de clasificación*. Si disponemos de un conjunto de entrenamiento con datos etiquetados.
  - Entrada: características de los recursos
  - Salida: Recomendable / No recomendable
- *Métodos de asociación*. Si no disponemos de etiquetas de clase
- Ambos métodos presentan el problema denominado del **arranque en frío** (*cold start*). “Al principio nunca se dispone de suficiente información para construir el modelo”.

Tareas en EDM

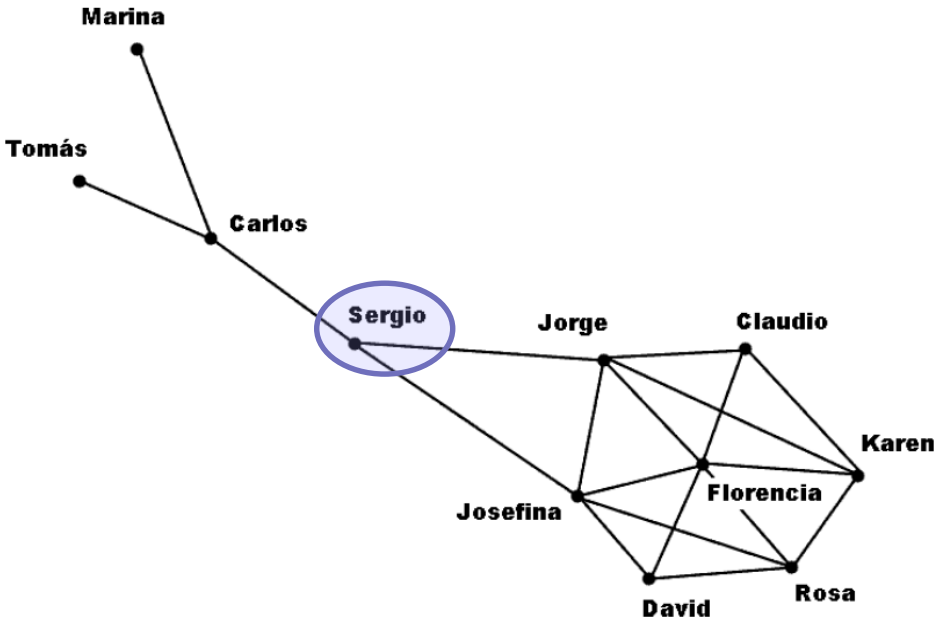
Tareas de alto nivel: Recomendar recursos o actividades a los estudiantes (III)

Filtrado colaborativo: Recommendamos a un usuario los mismos recursos que han funcionado bien en otros usuarios similares a éste.

- *Métodos de clustering*. Una vez que obtenemos los grupos de usuarios con características similares, podemos encontrar los recursos que éstos han utilizado y recomendarlos a nuevos usuarios que pertenezcan a ese grupo.
- Recientemente se ha empleado para esta tarea el *análisis de redes sociales*. En lugar de crear los grupos se recomiendan los recursos de los vecinos más cercanos en la red social.

Tareas en EDM

Tareas de alto nivel: Recomendar recursos o actividades a los estudiantes (IV)



En este ejemplo, se recomendarán a Sergio los recursos usados por Jorge, Josefina and Carlos

Tareas en EDM

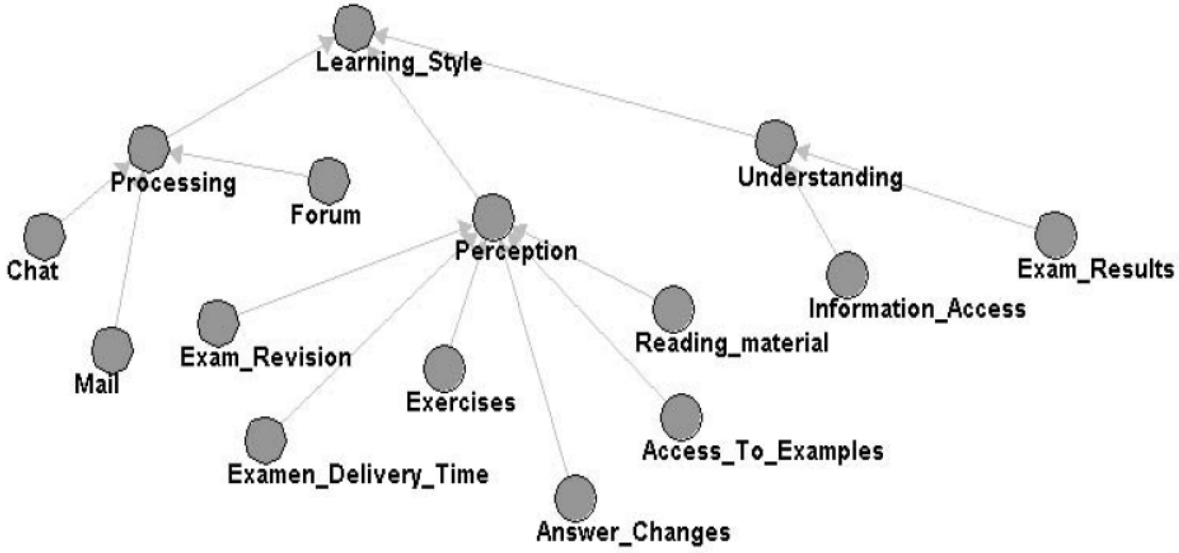
Tareas de alto nivel: Modelar el comportamiento del estudiante

- Consiste en desarrollar modelos cognitivos de los estudiantes, tanto a nivel de modelado de destrezas como de conocimiento declarativo.
- El interés de esta tarea es múltiple:
  - Permite la construcción de sistemas tutores inteligentes usando este modelo cognitivo y adaptarlos a las características del estudiante.
  - La información generada puede ayudar a comprender los mecanismos psicológicos que influyen en el aprendizaje.

Tareas en EDM

Tareas de alto nivel: Modelar el comportamiento del estudiante (II)

- Uno de los modelos más populares para representar el comportamiento de estudiantes son las redes bayesianas



- También se ha usado minería de reglas de asociación para modelar sistemas hipermedia adaptativos

Tareas en EDM

Tareas de alto nivel: Detección de comportamientos no deseados

- El concepto de comportamiento estudiantil no deseado es muy amplio, incluyendo:
  - Realización de acciones erróneas
  - Uso inapropiado de recursos
  - Intento de hacer trampas en el sistema
  - Otras cuestiones: detección de baja motivación, fracaso y abandono estudiantil.

Tareas en EDM

Tareas de alto nivel: Detección de comportamientos no deseados (II)

- *Clasificación*: Se construyen modelos para distinguir entre comportamiento deseado y no deseado [Bravo y Ortigosa, 2009; Dekker *et al.*, 2009]
- *Métodos de detección de anomalías*: Se aplican métodos de agrupamiento y se detectan datos que no pueden ser incluidos en ningún grupo [Burlak *et al.*, 2006; Vee *et al.*, 2006]
- *Minería de datos de asociación y descubrimiento de subgrupos*: Se encuentran para explicar el comportamiento anómalo de un grupo de estudiantes [Ma *et al.*, 2000; Hwang , 2005]

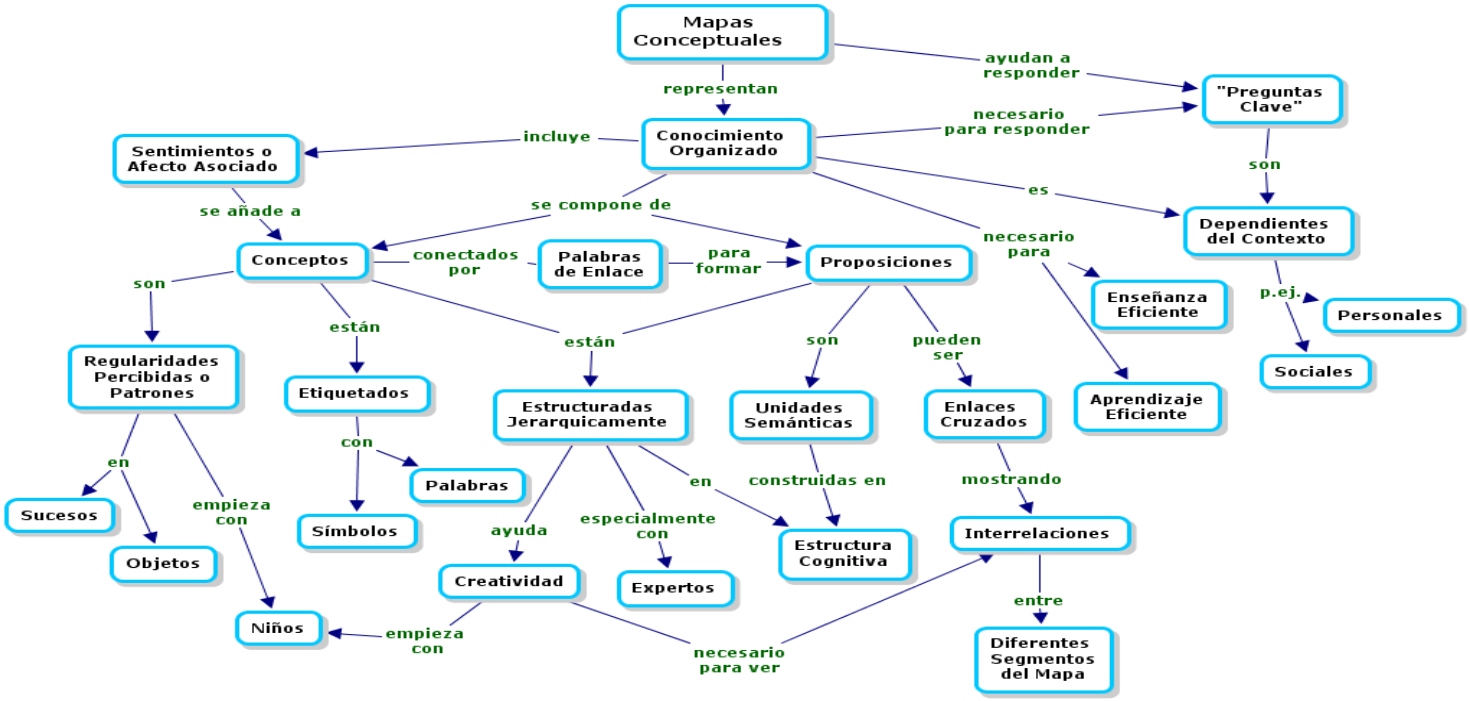
Tareas en EDM

Tareas de alto nivel: Generación automática de mapas conceptuales

- Los mapas conceptuales se utilizan para representar gráficamente conceptos o ideas que tienen una relación jerárquica.
- Un mapa conceptual es una red en la que los conceptos son los nodos de la red, y hay un número de aristas que sirven para relacionar unos conceptos con otros.
- Se trata de una forma estructurada de visualizar la información más relevante sobre un determinado tema.

Tareas en EDM

Tareas de alto nivel: Generación automática de mapas conceptuales (II)



El desarrollo de mapas conceptuales puede ser muy laborioso, especialmente cuando el dominio que se pretende representar el complicado.

Tareas en EDM

Tareas de alto nivel: Generación automática de mapas conceptuales (III)

- Se han utilizado dos tipos de técnicas para generar mapas conceptuales de forma automática:
  - *Minería de reglas de asociación*: Las reglas extraídas representan relaciones entre los conceptos que se incluirán en el mapa [Hwang, 2005; Tseng *et al.*, 2007; Lee *et al.*, 2009].
  - *Minería de textos*: Estas técnicas se han utilizado para extraer las palabras clave que representan los conceptos que se incluirán en el mapa [Chen et al., 2008]

Casos de éxito

Casos de éxito

Predicción de rendimiento académico

S. B. Kotsiantis

Use of machine learning techniques for educational proposes: a decision support system for forecasting students' grades

Artif Intell Rev (2012) 37:331–344

Casos de éxito

Predicción de rendimiento académico

- Modelo de regresión para predecir las calificaciones finales de 354 estudiantes de la *Hellenic Open University*, en la materia *Introduction to Informatics* (INF10).
- 17 attributos organizados en tres clases
  - Registry class
  - Tutor class
  - Classroom class
- Los mejores modelos presentan errors de predicción de un 12%

Student's Registry (demographic) attributes	
Sex	Male, female
Age	24–46
Marital status	Single, married, divorced, widowed
Number of children	None, one, two or more
Occupation	No, part-time, fulltime
Computer literacy	No, yes
Job associated with computers	No, junior-user, senior-user
Attributes from tutors' records	
1st Face to face meeting	Absent, present
1st Written assignment	No, 0–10
2nd Face to face meeting	Absent, present
2nd Written assignment	No, 0–10
3rd Face to face meeting	Absent, present
3rd Written assignment	No, 0–10
4th Face to face meeting	Absent, present
4th Written assignment	No, 0–10
Class	
Final examination test	0–10

Casos de éxito

Predicción de rendimiento académico – Variables más relevantes

Atribute	*Valor*	P_{apr}	P_{sus}
Assignment$_1$	Score > 3	0.02	0.19
	3 ≤ Score ≤ 6	0.14	0.35
	Score > 6	0.84	0.46
Assignment$_2$	Score > 3	0.08	0.52
	3 ≤ Score ≤ 6	0.15	0.26
	Score > 6	0.77	0.22
Assignment$_3$	Score > 3	0.03	0.61
	3 ≤ Score ≤ 6	0.21	0.20
	Score > 6	0.66	0.19
Assignment$_4$	Score > 3	0.04	0.68
	3 ≤ Score ≤ 6	0.31	0.15
	Score > 6	0.65	0.17
Interview$_2$	Present	0.22	0.54
	Absent	0.78	0.46
Interview$_3$	Present	0.20	0.65
	Absent	0.80	0.35
Intgerview$_4$	Present	0.23	0.76
	Absent	0.77	0.24

Las variables más relevantes son las denominadas “Experience in Computing” (51%) y “Previous Work” (52%)

Casos de éxito

Detección precoz del fracaso escolar

M.A. Jiménez, J.M. Luna, S. Ventura

EDM para la detección precoz del fracaso escolar en Secundaria

Taller de Minería de Datos y Aprendizaje (TAMIDA 2013). Madrid, 2013

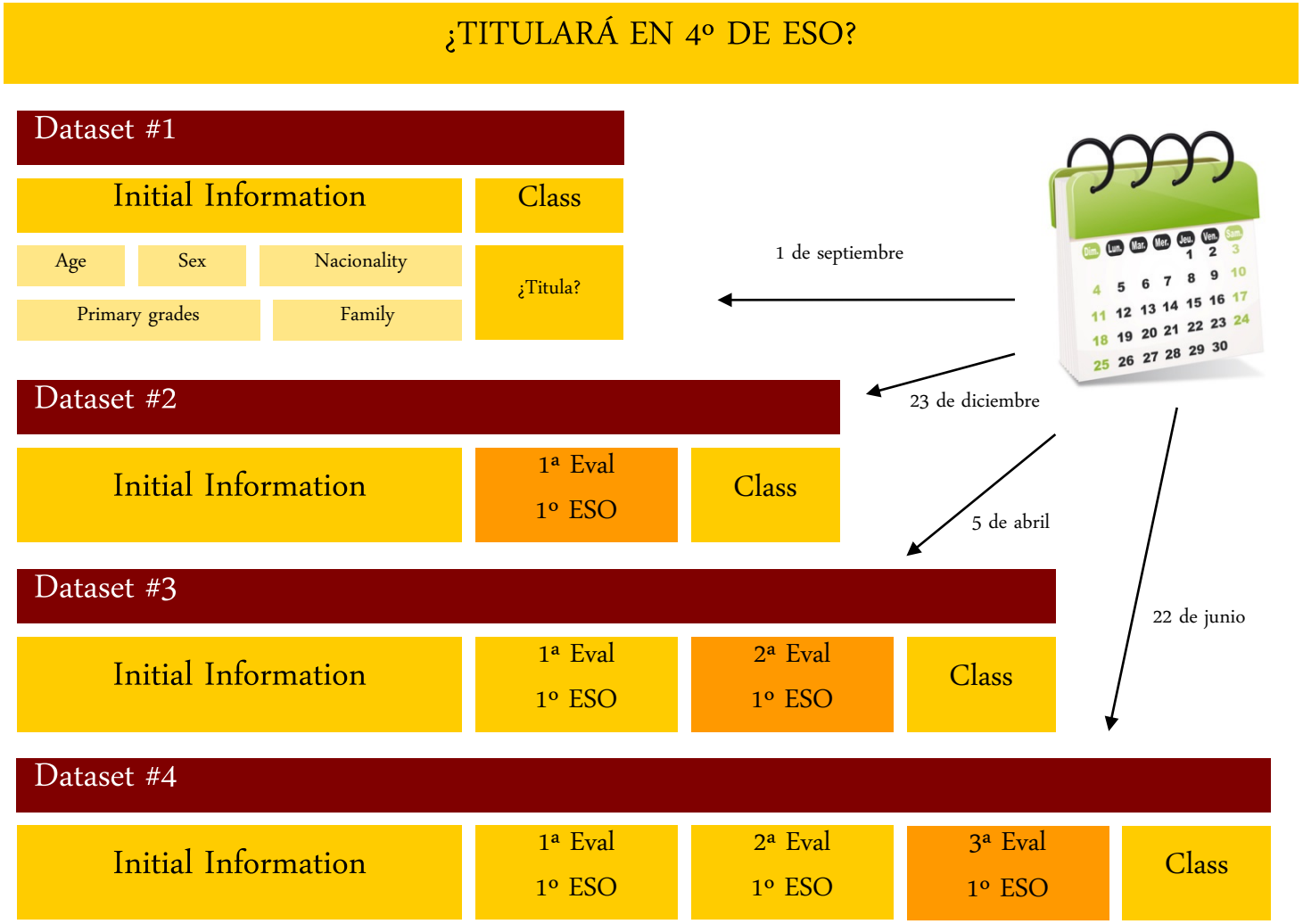
Casos de éxito

Detección precoz del fracaso escolar

- Hasta la fecha, la mayoría de los trabajos sobre predicción del rendimiento académico usaban información de interacción con el alumno en un curso.
- Nuestro propósito en este caso era usar la información disponible en los registros institucionales:
  - Información demográfica, social
  - Información académica (datos históricos y resultados parciales)
- También nos interesaba realizar un estudio sobre el rango de tiempo necesario para obtener buenos modelos predictivos.
- Entorno académico: Educación secundaria obligatoria

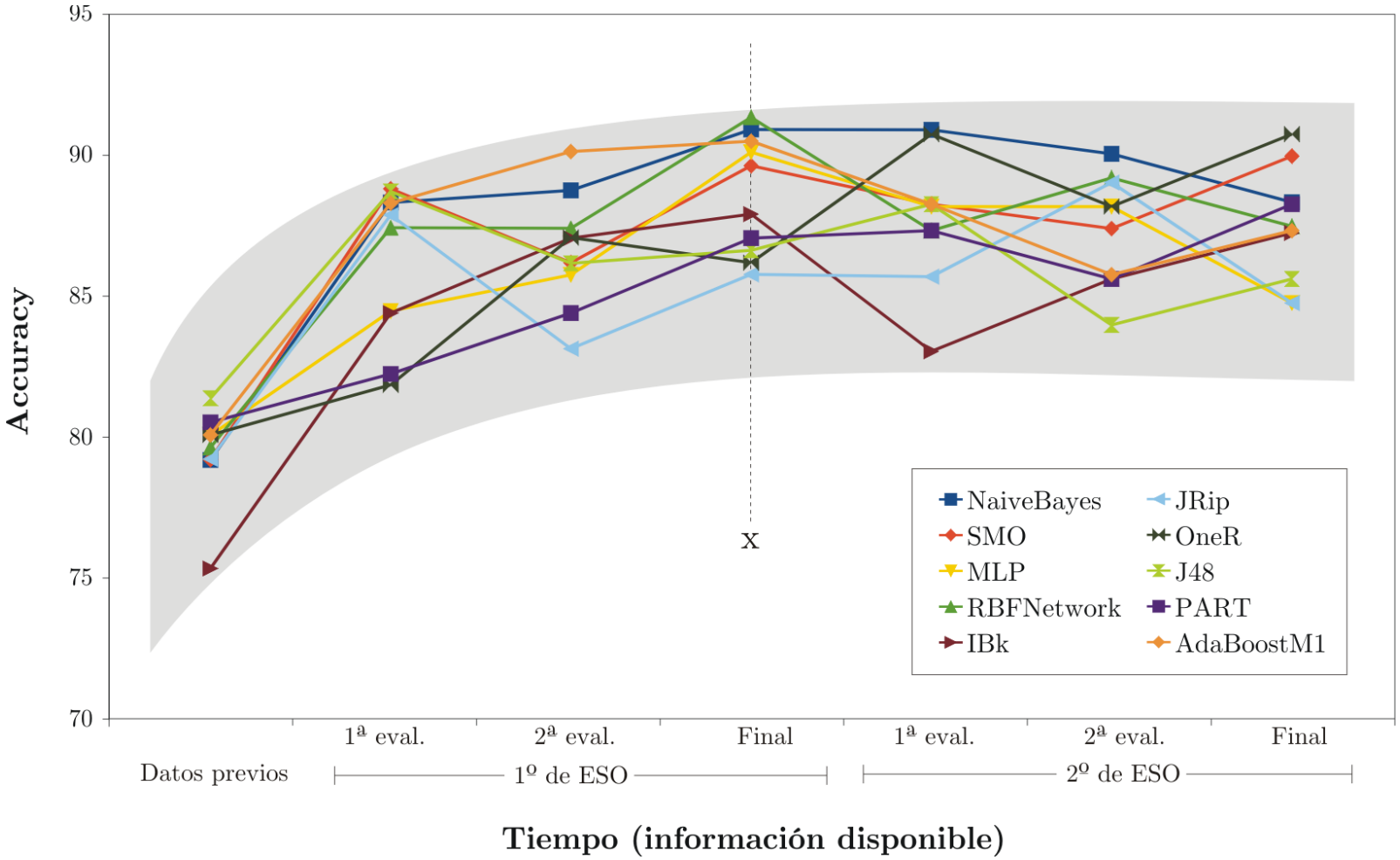
Casos de éxito

Detección precoz del fracaso escolar



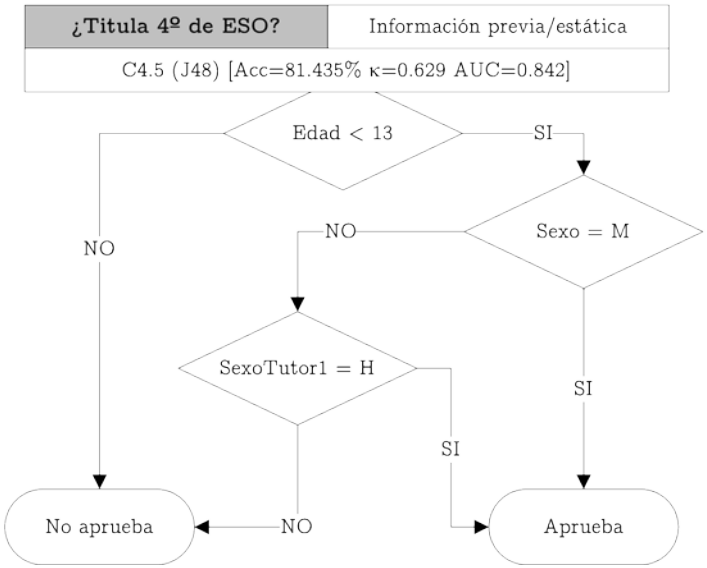
Casos de éxito

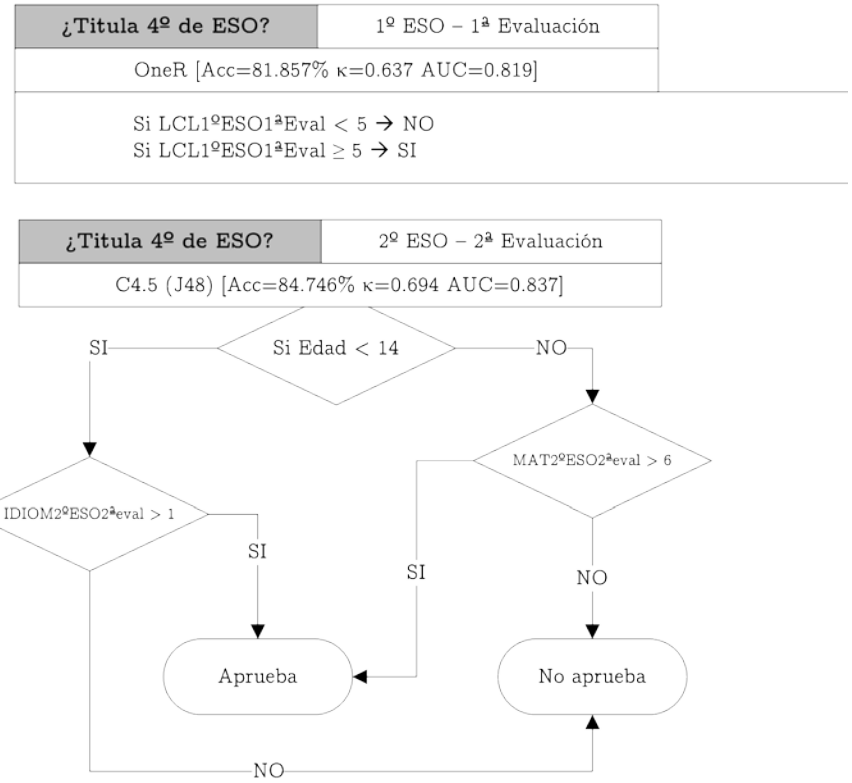
Detección precoz del fracaso escolar



Casos de éxito

Detección precoz del fracaso escolar





Algunos modelos descubiertos

Casos de éxito

Detección precoz del fracaso escolar

Cuestiones abiertas

- ¿Cómo de útil es el conocimiento descubierto?
  - Modificar la estrategia pedagógica empleada hasta la fecha y comprobar los resultados obtenidos
- Cómo extender los modelos obtenidos a varios centros educativos en una misma zona, region,…
  - Agrupamiento de centros
  - Datos contextuales

Casos de éxito

Extracción de información relevante mediante el análisis de redes sociales

R. Rabbany, M.Takaffoli & O. R. Zaiane

Social Network Analysis and Mining to Support the Assessment of Online Student Participation

ACM SIGKDD Explorations, December 2011

Casos de éxito

Extracción de información relevante mediante el análisis de redes sociales

- Meerkat-ED es una herramienta desarrollada en el Center for Innovation on Machine Learning (AICML) de la Universidad de Alberta.
- La herramienta aplica el análisis de redes sociales sobre el contenido de los mensajes intercambiados en una serie de foros educativos.
- La visualización de estas estruturas puede ayudar a los profesores a comprender la participación de los estudiantes en las discusiones *on-line*.

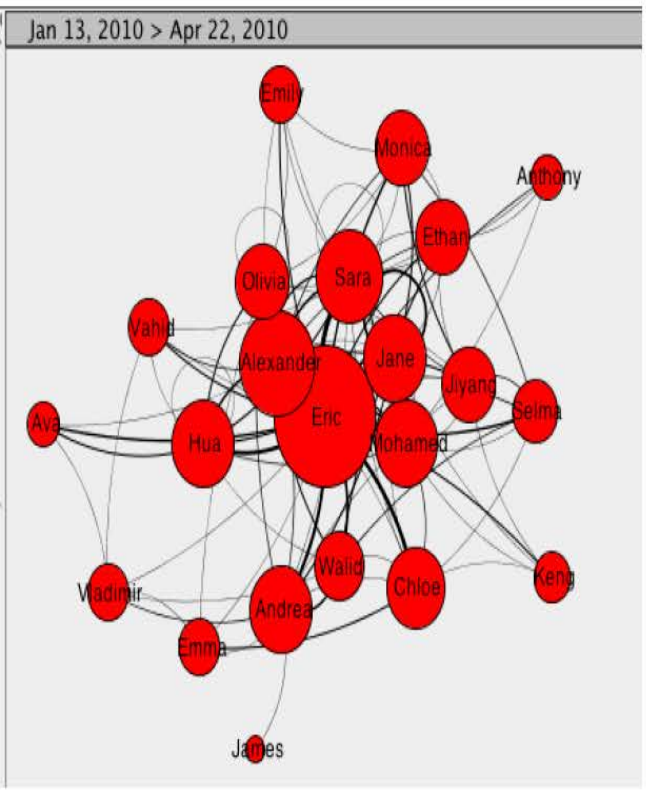
Casos de éxito

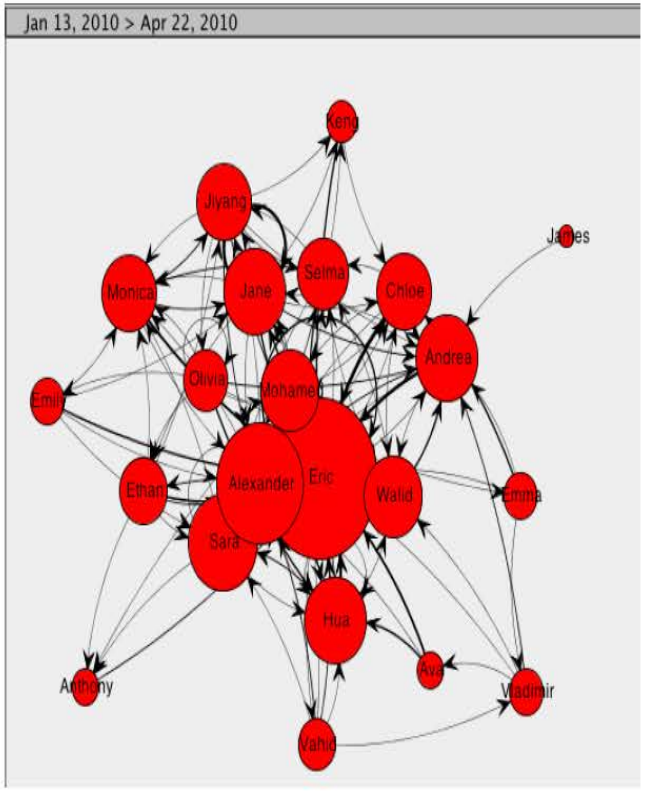
Extracción de información relevante mediante el análisis de redes sociales

- Meerkat-ED prepara instantáneas de los participantes en los foros de discusión, sus interacciones y diferencia entre estudiantes centrales y periféricos.
- Esto da al profesor una visión rápida de lo que está pasando en los foros de discusión.
- La herramienta muestra también cómo participa cada estudiante en cada tema, informando sobre su centralidad en la discusión de cada ítem, el número de envíos y los términos usados por este estudiante en la discusión.

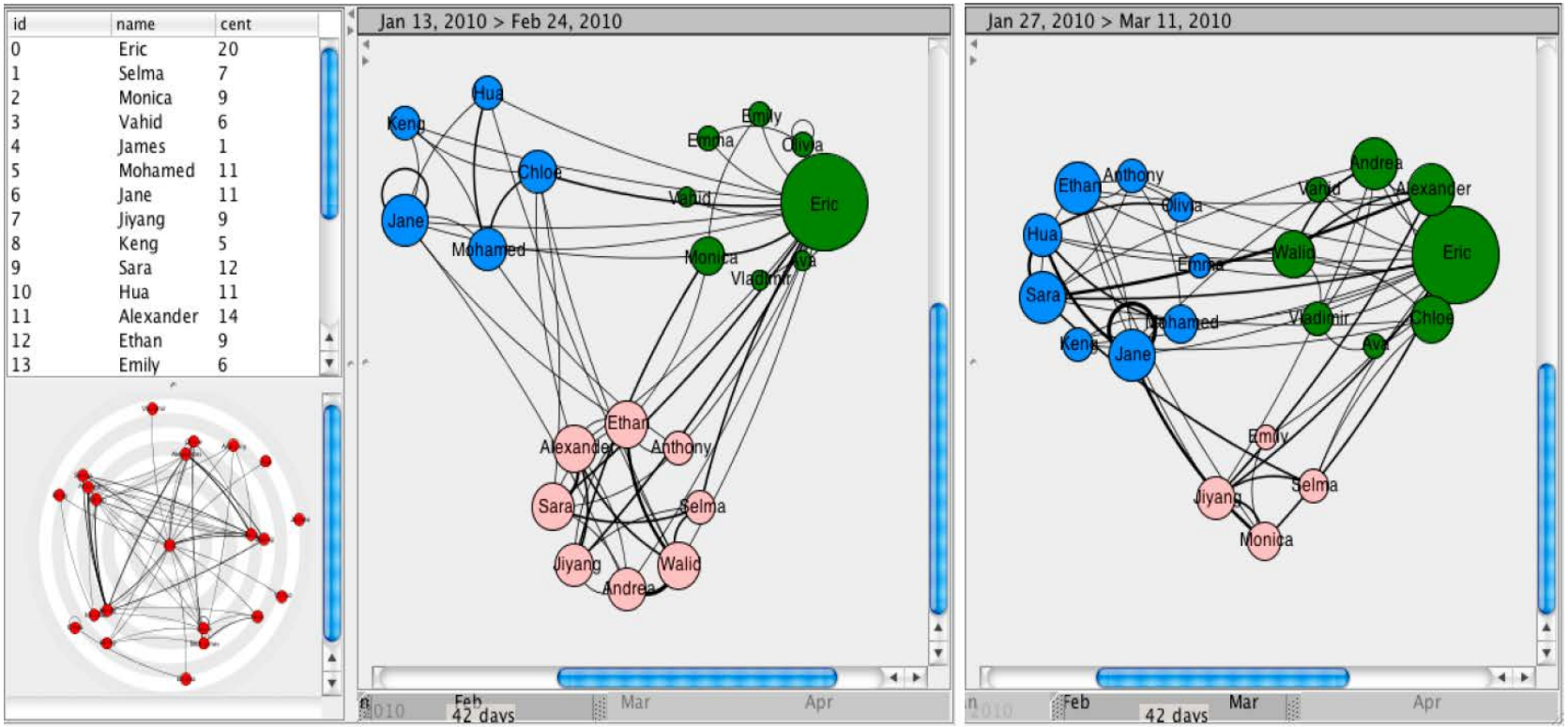
Providing Feedback based on Social Networks Analysis

id	name	cent
0	Eric	20
1	Selma	7
2	Monica	9
3	Vahid	6
4	James	1
5	Mohamed	11
6	Jane	11
7	Jiyang	9
8	Keng	5
9	Sara	12
10	Hua	11
11	Alexander	14
12	Ethan	9
13	Emily	6
14	Anthony	4
15	Olivia	9
17	Chloe	10
16	Ava	4
19	Vladimir	6
18	Walid	8
21	Emma	6
20	Andrea	11





Providing Feedback based on Social Networks Analysis



Groups of students who collaborate over time (teacher can easily monitor the progress of these groups over time).

Nuevos retos / Problemas abiertos

Nuevos retos / Problemas abiertos

- Desarrollo de herramientas específicas para EDM.
  - La mayoría de los trabajos publicados utilizan herramientas estándar o implementaciones de los autores
  - Ya están apareciendo algunas plataformas, pero son demasiado específicas para un determinado problema/entorno.
- Integración con sistemas de e-learning o con plataformas de gestión académica.
  - La falta de integración existente en la actualidad obliga un proceso de recopilación y preparación de datos que suele ser muy costoso
  - Los sistemas desarrollados hasta la fecha suelen estar orientados a un único entorno, y su aplicación a otros problemas es compleja

Nuevos retos / Problemas abiertos

- Estandarización de métodos y datos.
  - En la actualidad no existen formatos estándar para la representación de datos educativos.
  - Se trata de un problema complejo, porque existe una gran variedad de datos.
  - Hay un par de intentos por parte de algunas organizaciones pero hasta ahora han sido infructuosos
  - También se echa de en falta el desarrollo de guías de buenas prácticas para la resolución de tareas en EDM
- Validación de métodos e incorporación al proceso de enseñanza-aprendizaje

Nuevos retos / Problemas abiertos

Mejorando MOOCs con EDS

- Los MOOCs (Massive Open Online Courses) son sistemas de aprendizaje online de libre acceso que suele presentar un gran número de alumnos matriculados (más de 10000 por curso).
  - Ejemplo: MOOC sobre Inteligencia Artificial de S. Thrun y P. Norvig contó en 2011 con 160000 alumnos inscritos.
- Los MOOCs plantean múltiples retos que pueden ser abordados desde la perspectiva de Educational Data Science (o de EDM, si mantenemos la terminología clásica).
  - Algunos son problemas ya conocidos, pero agravados por el gran volumen de información disponible (Big Data)
  - Otros son problemas totalmente nuevos.

Nuevos retos / Problemas abiertos

Mejorando MOOCs con EDS (II)

- Modelado del comportamiento estudiantil y de las interacciones con el sistema
  - Los profesores en un MOOC necesitan herramientas que les ayuden a tutorizar mejor a los estudiantes, evaluar qué recursos son los más apropiados / exitosos, etc.
- Predicción de abandono:
  - En media los MOOCs presenta ratios de abandono de un 90%.
  - Se necesitan herramientas que ayuden a predecir este abandono y den soporte a políticas de captación de alumnos.

Nuevos retos / Problemas abiertos

Mejorando MOOCs con EDS (III)

- Evaluación y retroalimentación
  - Un tema muy común en los MOOCs (sobre todo en disciplinas no científicas) es la evaluación por pares o la autoevaluación.
  - Se ha comprobado que ambos sistemas suelen diferir de las calificaciones proporcionadas por el profesor.
  - Se necesitan herramientas que ayuden a modelar o refinar estos procesos de evaluación.
- Personalización del entorno para estudiantes
  - El ideal del aprendizaje online es la personalización. El sistema debería encontrar los recursos mejores para cada estudiante y presentárselos para optimizar su rendimiento en la plataforma.
  - El problema es análogo a los que se han planteado en EDM, con la complicación asociada a volumen de datos disponible

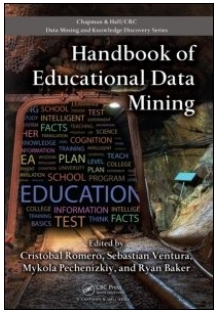
Recursos interesantes

Recursos interesantes

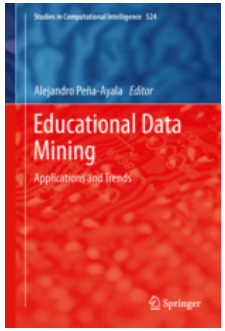
Recursos bibliográficos

- *Data Mining in E-Learning*.
  C. Romero & S. Ventura (Eds).
  Editorial WIT Press, 2006.

- Handbook of Educational Data Mining.
  C.Romero, S. Ventura, M. Pechenizky & R. Baker (Eds).
  Editorial CRC Press, Taylor & Francis Group. 2010.



- *Education Data Mining: Applications and Trends*.
  A. Peña-Ayala (Eds).
  Springer, SCI Vol. 524, 2014



Recursos interesantes

Recursos bibliográficos – Artículos de revisión

- C. Romero & S. Ventura. Educational Data Mining: A survey from 1995 to 2005. *Expert Systems with Applications* 33:1, pp. 135-146, 2007.
- Castro, F., Vellido, A., Nebot, A. Mugica, F. Applying Data Mining Techniques to e-Learning Problems. In: Evolution of Teaching and Learning Paradigms in Intelligent Environment. Studies in Computational Intelligence, 62, Springer-Verlag, 183-221. 2007.
- Baker, R., Yacef, K. The State of Educational Data Mining in 2009: A Review and Future Visions. Journal of Educational Data Mining, 1, 1, 3-17. 2009.
- Peña, A. Dominguez, R. Medel, J.J. Educational data mining: a sample of review and study case. World Journal on Educational Technology, 2, 118-139. 2009.

Recursos interesantes

Recursos bibliográficos – Artículos de revisión

- C. Romero, S. Ventura. Educational Data Mining: A Review of the State-of-the-Art. IEEE Transactions on Systems, Man, and Cybernetics--Part C: Applications and Reviews. 40:6, pp. 601 – 618. 2010.
- Baker, R.S.J.d. Data Mining for Education. In McGaw, B., Peterson, P., Baker, E. (Eds.) International Encyclopedia of Education (3rd edition), vol. 7, pp. 112-118. Oxford, UK: Elsevier. 2010.
- Scheuer, O. & McLaren, B.M. Educational Data Mining. In the Encyclopedia of the Sciences of Learning, Springer. 2011.
- Karen Cator. Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics. Report of the U.S. Office of Educational Technology. 2012.
- C. Romero, S. Ventura. Data Mining in Education. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. Volume 3, Issue 1, pages 12–27, 2013.

Recursos interesantes

Recursos bibliográficos – Artículos de revisión

- R. Jindal, M. D. Borah. A survey on educational data mining and research trends. International Journal of Database Management Systems. Vol.5, No.3, June 2013.
- Al-Razgan, M., Al-Khalifa, A., Al-Khalifa, H. Educational Data Mining: A systematic review of the published literature 2006-2013. Int. Conference on Advanced Data and Information Engineering. 711-719. 2014.
- Peña-Ayala, A. Educational data mining: A survey and a data mining-based analysis of recent works. Expert Systems with Applications. 41, 1432-1462. 2014.
- Baker, R.S. Educational Data Mining: An Advance for Intelligent Systems in Education. IEEE Inteligent Systems. 78-82, 2014.

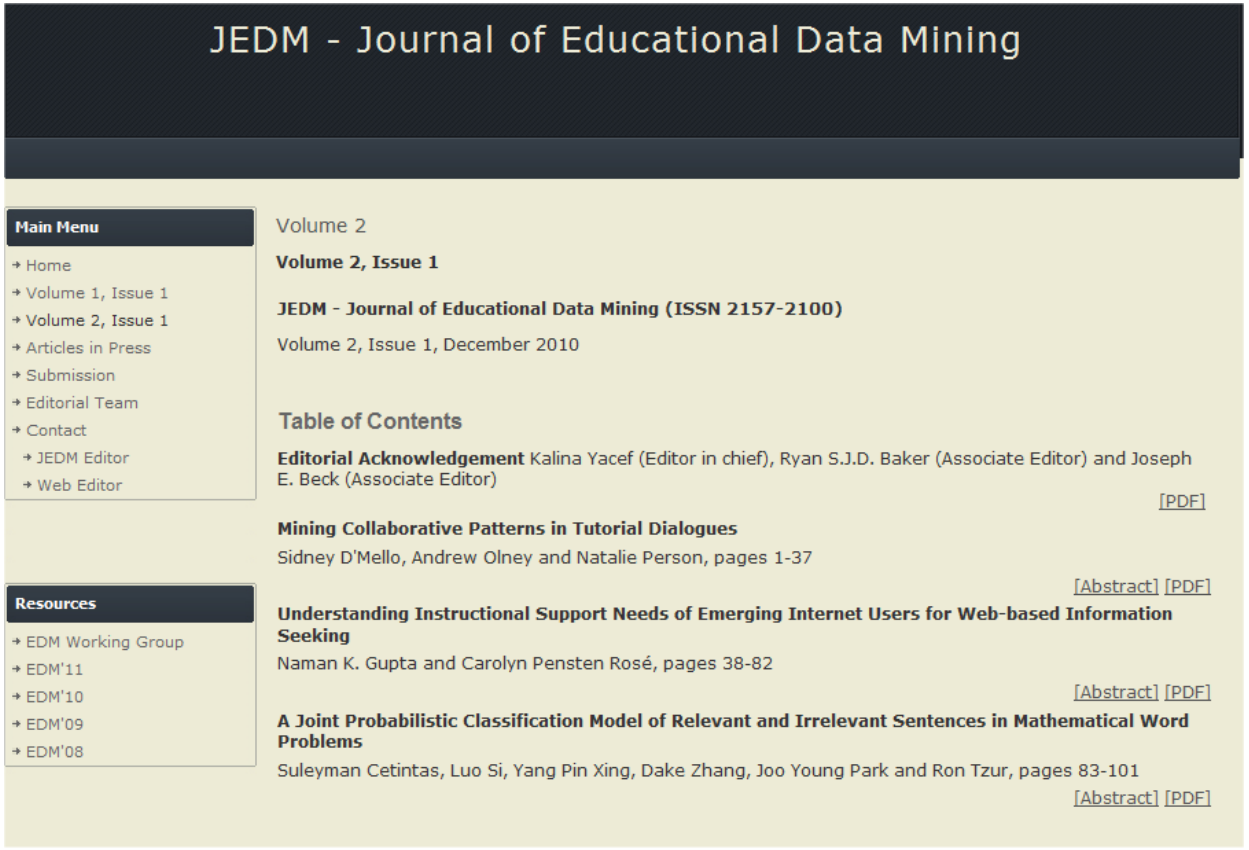
Recursos interesantes

Educational Data Mining Society

http://www.educationaldatamining.org/

Recursos interesantes

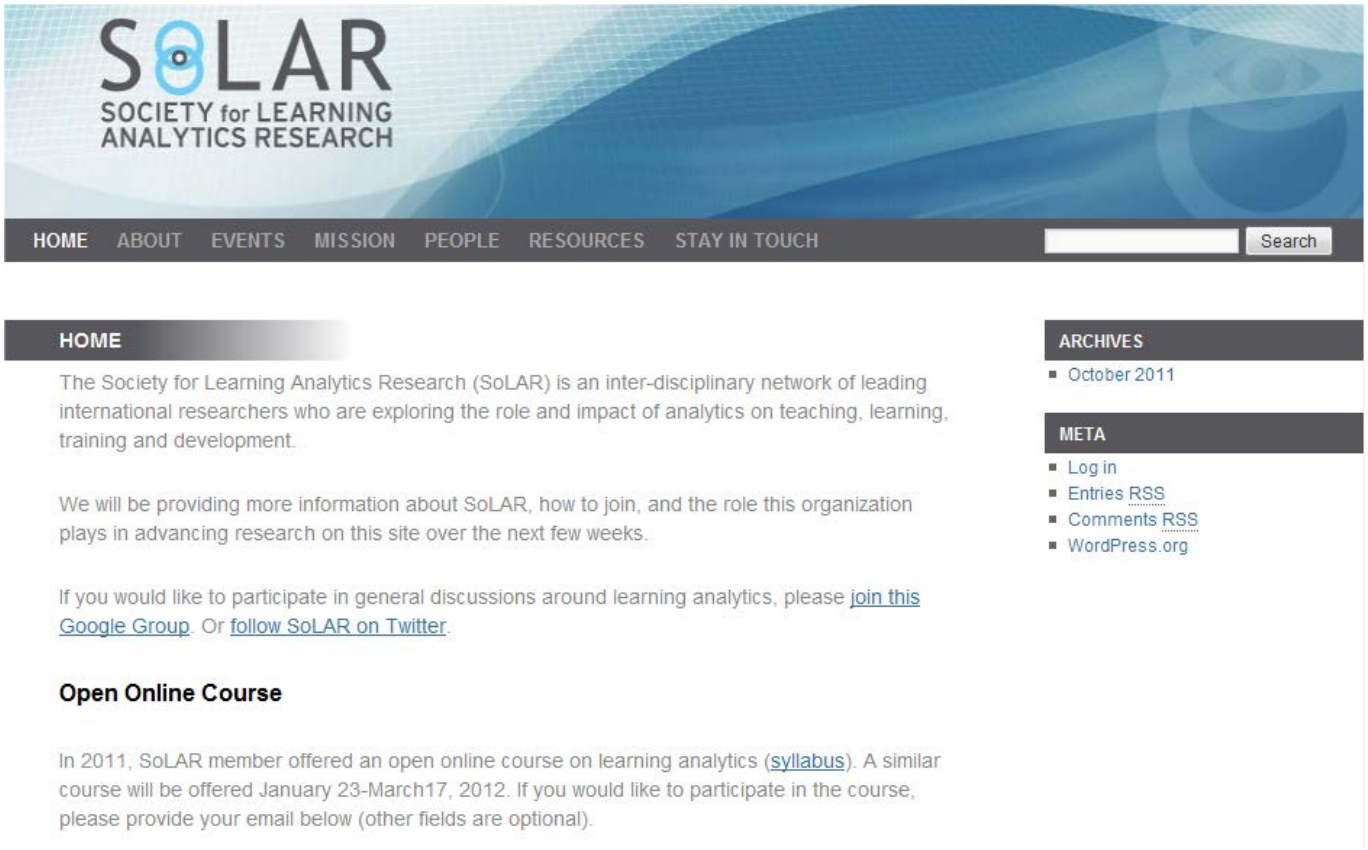
Journal of Educational Data Mining



http://www.educationaldatamining.org/JEDM/

Recursos interesantes

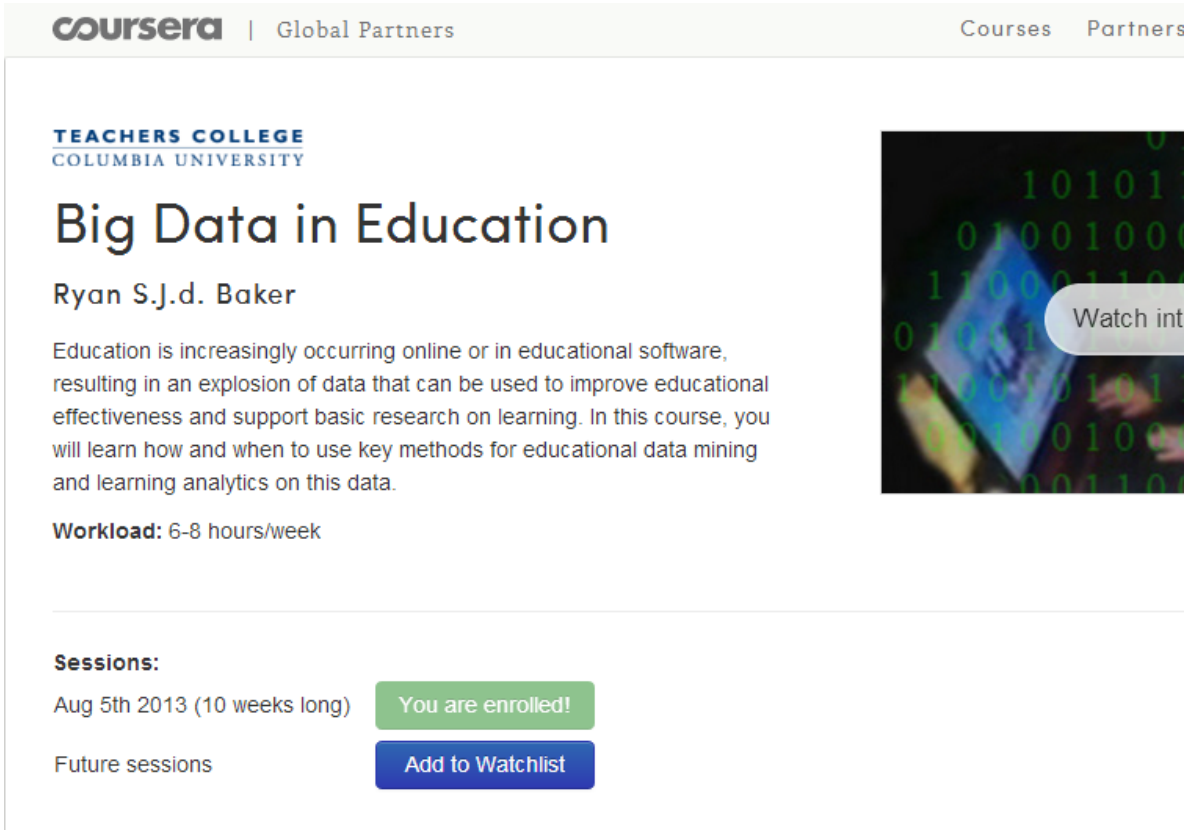
Society for Learning Analytics



http://www.solaresearch.org/

Recursos interesantes

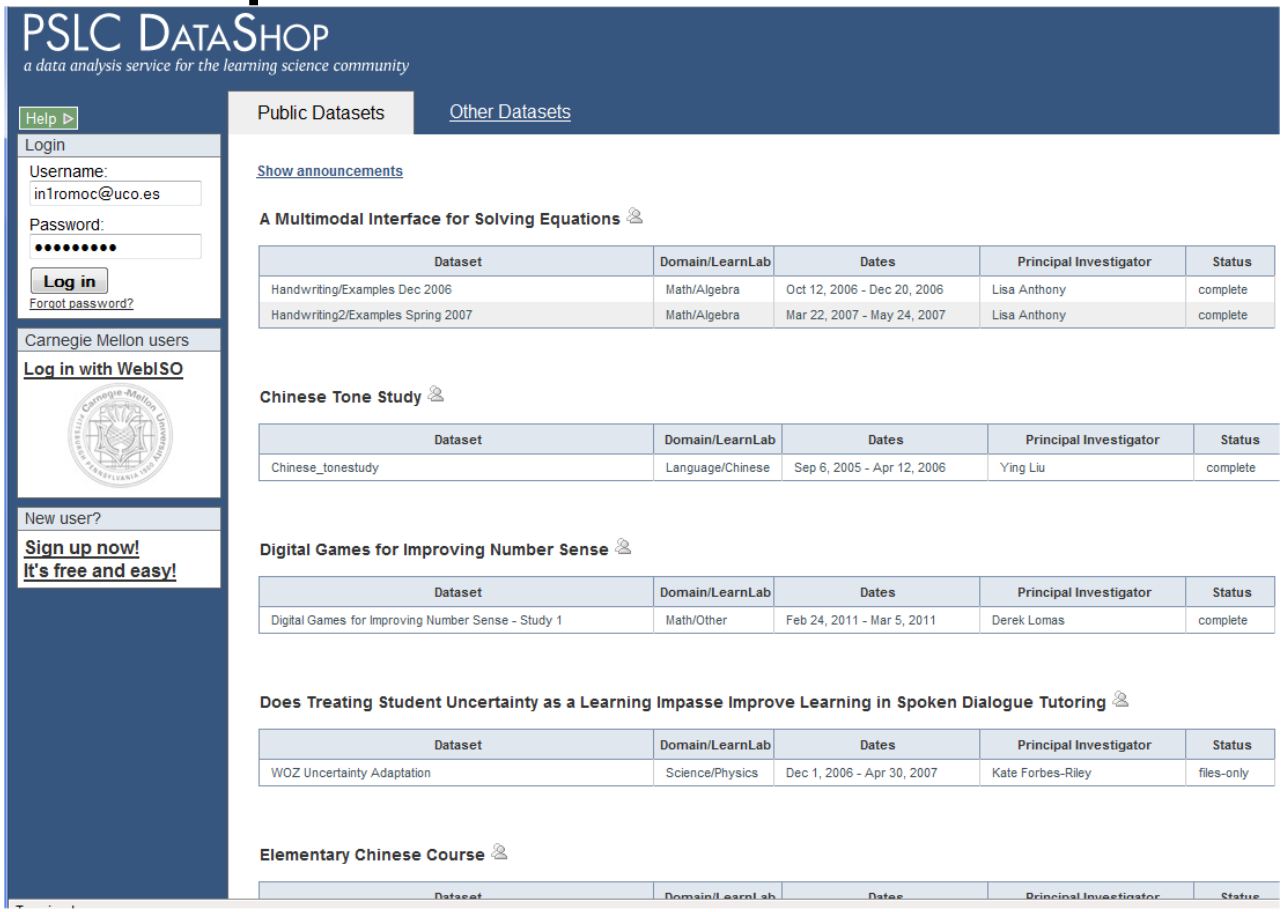
Coursera: Big Data in Education



https://www.coursera.org/course/bigdata-edu

Recursos interesantes

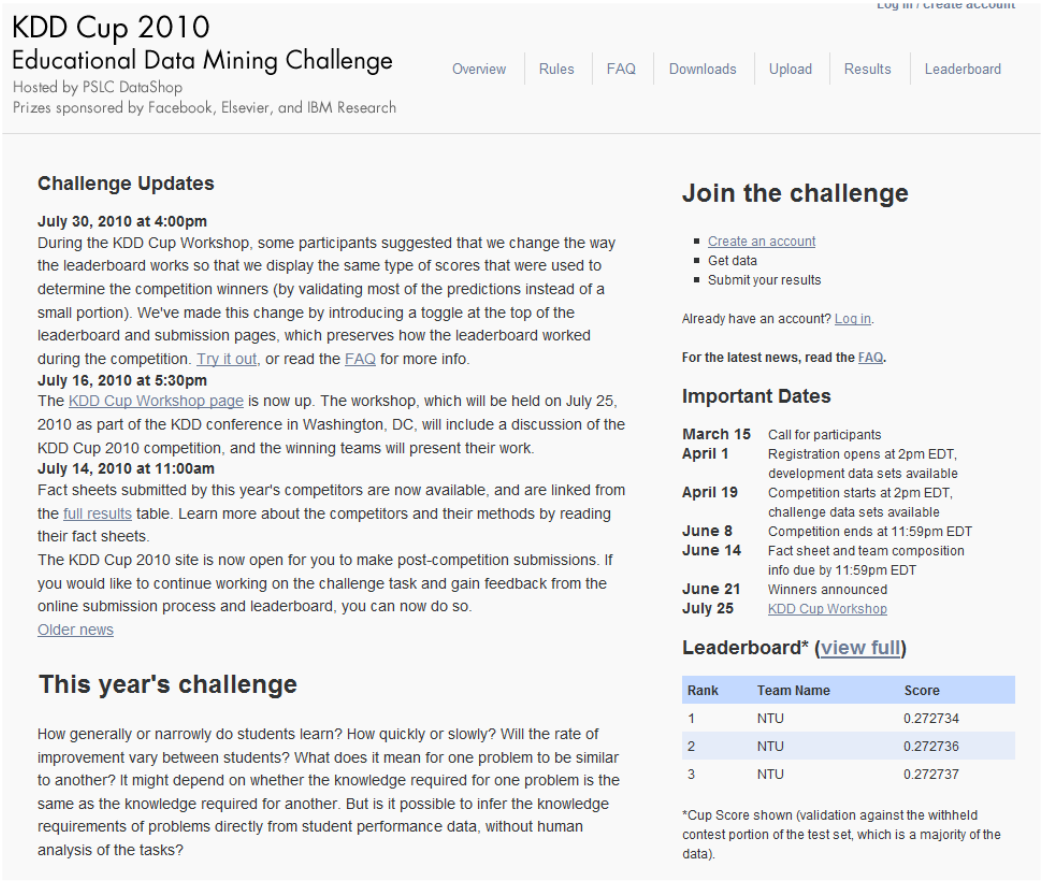
PSLC Data Shop



https://pslcdatashop.web.cmu.edu/

Recursos interesantes

KDD Cup 2010



KDD Cup 2010

Educational Data Mining Challenge

Hosted by PSLC DataShop

Prizes sponsored by Facebook, Elsevier, and IBM Research

Overview | Rules | FAQ | Downloads | Upload | Results | Leaderboard

Challenge Updates

July 30, 2010 at 4:00pm

During the KDD Cup Workshop, some participants suggested that we change the way the leaderboard works so that we display the same type of scores that were used to determine the competition winners (by validating most of the predictions instead of a small portion). We've made this change by introducing a toggle at the top of the leaderboard and submission pages, which preserves how the leaderboard worked during the competition. Try it out, or read the FAQ for more info.

July 16, 2010 at 5:30pm

The KDD Cup Workshop page is now up. The workshop, which will be held on July 25, 2010 as part of the KDD conference in Washington, DC, will include a discussion of the KDD Cup 2010 competition, and the winning teams will present their work.

July 14, 2010 at 11:00am

Fact sheets submitted by this year's competitors are now available, and are linked from the full results table. Learn more about the competitors and their methods by reading their fact sheets.

The KDD Cup 2010 site is now open for you to make post-competition submissions. If you would like to continue working on the challenge task and gain feedback from the online submission process and leaderboard, you can now do so.

Older news

This year's challenge

How generally or narrowly do students learn? How quickly or slowly? Will the rate of improvement vary between students? What does it mean for one problem to be similar to another? It might depend on whether the knowledge required for one problem is the same as the knowledge required for another. But is it possible to infer the knowledge requirements of problems directly from student performance data, without human analysis of the tasks?

Join the challenge

- Create an account
- Get data
- Submit your results

Already have an account? Log in.

For the latest news, read the FAQ.

Important Dates

March 15	Call for participants
April 1	Registration opens at 2pm EDT, development data sets available
April 19	Competition starts at 2pm EDT, challenge data sets available
June 8	Competition ends at 11:59pm EDT
June 14	Fact sheet and team composition info due by 11:59pm EDT
June 21	Winners announced
July 25	KDD Cup Workshop

Leaderboard* (view full)

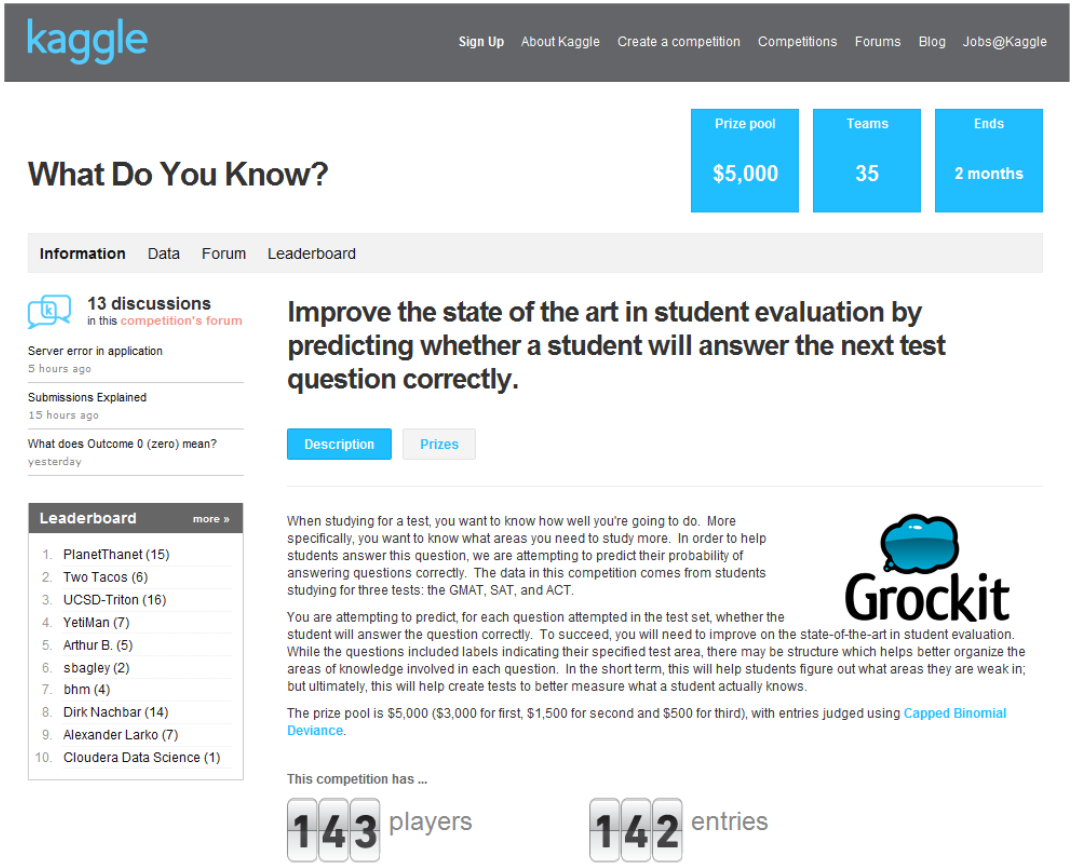
Rank	Team Name	Score
1	NTU	0.272734
2	NTU	0.272736
3	NTU	0.272737

*Cup Score shown (validation against the withheld contest portion of the test set, which is a majority of the data).

https://pslcdatashop.web.cmu.edu/KDDCup/

Recursos interesantes

Kaggle competition



http://www.kaggle.com/c/WhatDoYouKnow

Gracias por vuestra atención

¿Preguntas?

Sebastián Ventura

Department of Computer Sciences and Numerical Analysis. University pf Córdoba
Department of Computer Science. King Abdulaziz University