Ciencia de Datos Educativos

Sebastián Ventura Soto

Grupo de Investigación *Knowledge Discovery and Intelligent Systems*

Universidad de Córdoba

KDIS lab

DaSCI Instituto Andaluz Interuniversitario en Data Science and Computational Intelligence

UNIVERSIDAD DE CÓRDOBA

Contenidos

- Introducción
- Ciencia de Datos Educativos
  - Perspectiva Histórica
  - Disciplinas asociadas
  - Ciclo de vida en CDE
- Casos de Éxito
  - Predicción del rendimiento académico
  - Personalización de la enseñanza
  - Formación de grupos y/o equipos de trabajo
  - Evaluación por pares y autoevaluación
  - Recomendación de itinerarios
- Nuevos retos
  - Problemas abiertos
- Conclusiones

Introducción

Datos, datos... y más datos

- Vivimos en la era de los datos
- Los datos han cobrado una enorme importancia.
- La explotación de estos datos es aun más importante. Entre otras cosas permite:
  - Descubrir nuevos patrones
  - Comprobar/validar teorías científicas
  - Construir modelos para predecir el comportamiento de un sistema
  - Categorizar objetos
  - ...

Información generada en los Sistemas Educativos

- Instituciones educativas
  - Información de los estudiantes
  - Registros académicos
- Educación presencial
  - Múltiples datos generados por el alumno y su interacción con el profesor
  - Tamaño pequeño
- Sistemas tutores inteligentes
  - Acciones realizadas por los estudiantes
  - Resultados de la evaluación
  - ...
- Otros sistemas on-line (incluidos MOOCs)
  - Interacción con las plataformas de educación:
    - Páginas visitadas
    - Actividades realizadas
    - Comunicación con otros estudiantes
  - Resultados de la evaluación
- Simuladores y/o sistemas basados en gamificación
  - Acciones realizadas en el sistema
  - Progreso del aprendiz
  - Caminos recorridos
  - ...

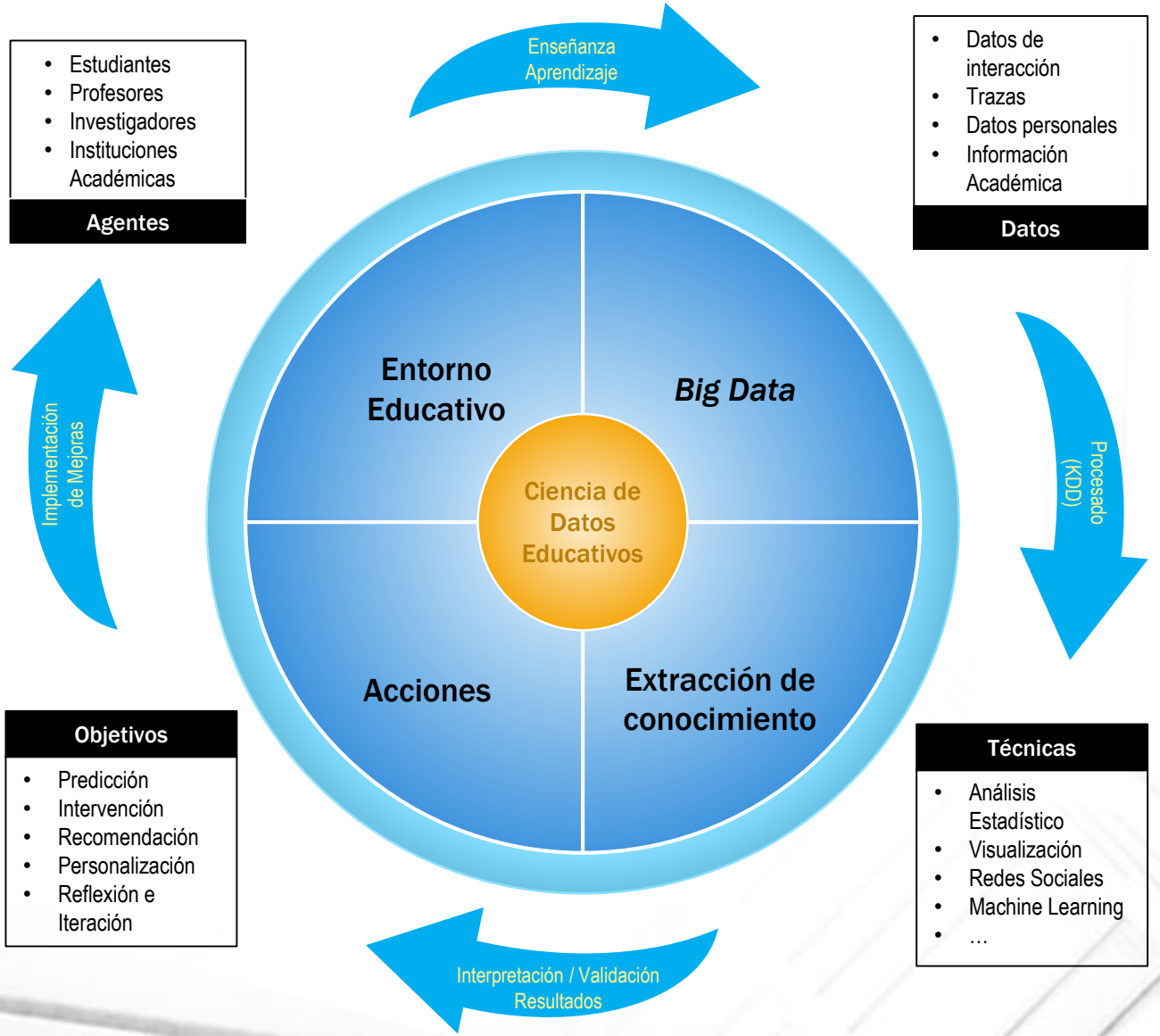
Extracción de conocimiento en los Sistemas Educativos

Toda la información generada por los sistemas educativos puede utilizarse en la extracción de nuevo conocimiento útil

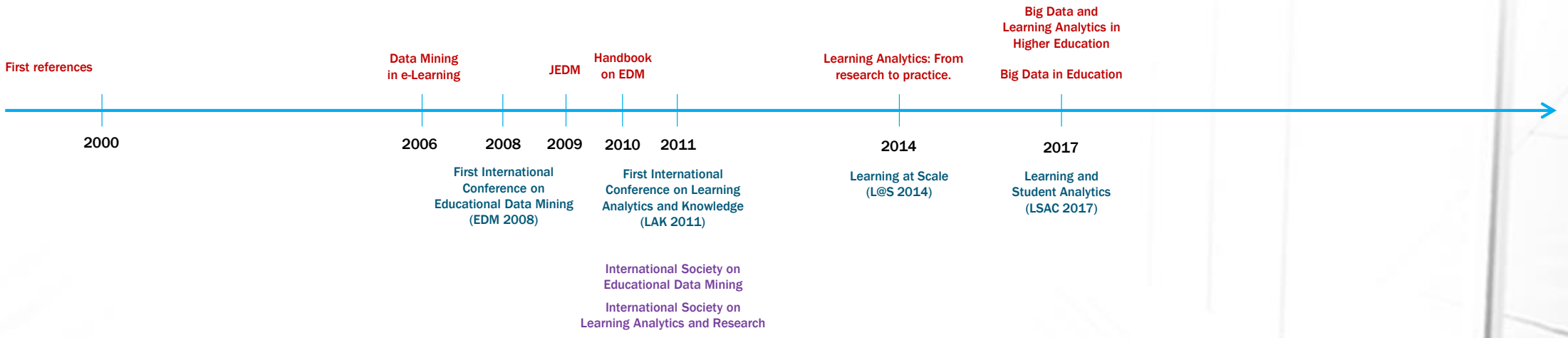
El término **Ciencia de Datos Educativos** (*Educational Data Science*) abarca, de forma genérica, a multitud de disciplinas cuyo objetivo es aplicar las distintas técnicas de la Ciencia de Datos sobre la información generada en los sistemas educativos.

Ciencia de Datos Educativos

Ciclo de Vida en EDS



Ciencia de Datos Educativos. Una perspectiva histórica

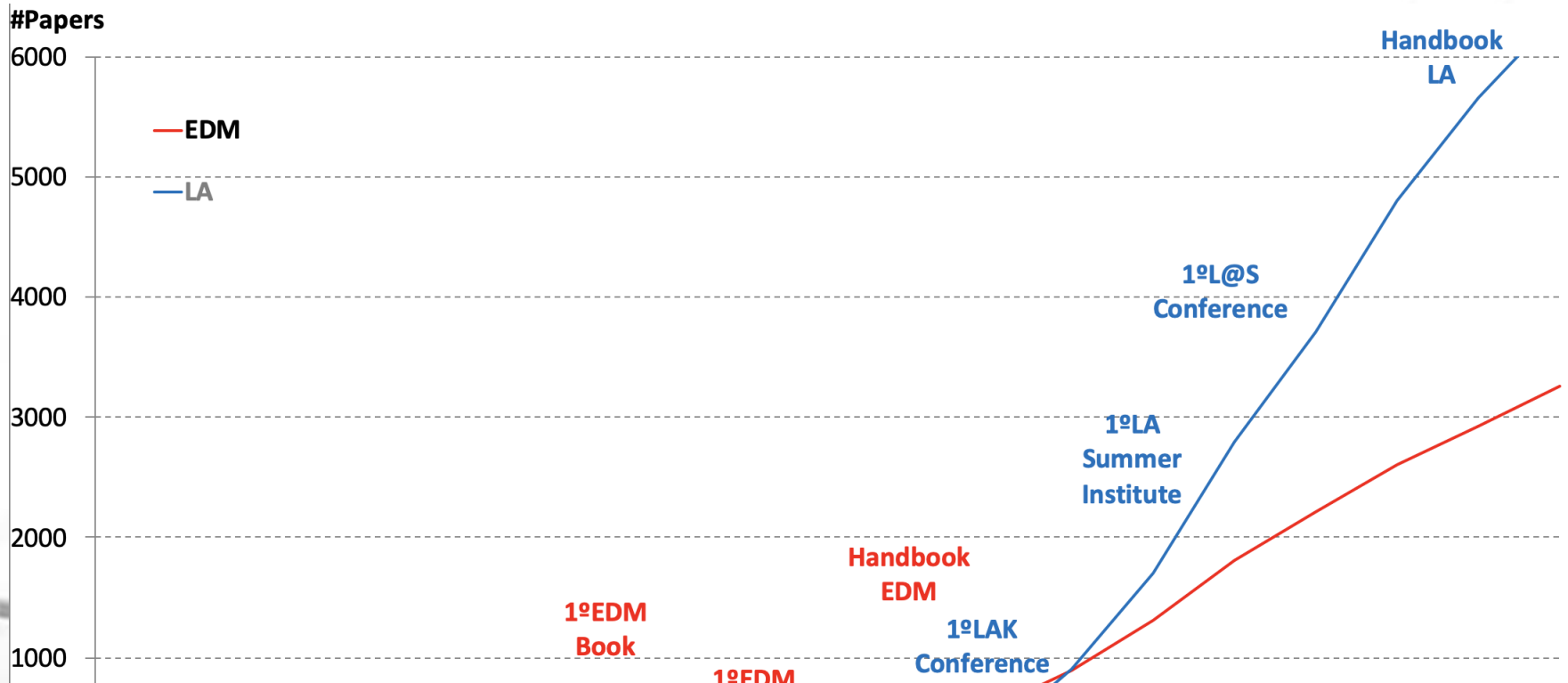


Ciencia de Datos Educativos: Disciplinas... ¿Incluidas?

El término Ciencia de Datos Educativos no está demasiado extendido. De hecho, existen términos mucho más populares que se refieren a la misma o a disciplinas parecidas

- **Minería de Datos Educativos** (*Educational Data Mining*, EDM). Aplicación de las técnicas de extracción de conocimiento a la mejora de los sistemas educativos.
- **Analítica de Aprendizaje** (*Learning Analytics*, LA). Aplicación de técnicas de analítica de datos para la mejora del proceso de enseñanza-aprendizaje.
- **Analítica Académica** (*Academic Analytics*, AA). Aplicación de las técnicas de analítica de datos e inteligencia de negocio a datos de las instituciones académicas.
- **Analítica de la Enseñanza** (*Teaching analytics*, TA). Analisis de datos de enseñanza y rendimiento académico.
- **Big Data en Educación** (Big Data in Education, BDE). Aplicación de técnicas de extracción de conocimiento en bases de datos educativas masivas, como las que se generan en los MOOCs

Ciencia de Datos Educativos. Evolución del Número de Publicaciones



Casos de éxito

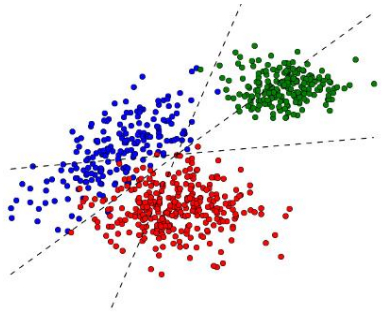
Predicción del rendimiento académico
Concepto

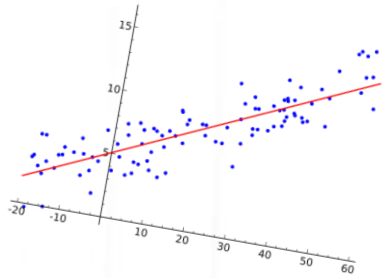
- Es una de las aplicaciones más estudiadas
- Se trata de obtener modelos para predecir el rendimiento del estudiante a partir de datos históricos
- Objetivos:
  - Detectar fracaso y/o abandono escolar
  - Mejor comprensión de ambos fenómenos.
  - Desarrollar medidas correctoras basadas en el conocimiento extraído de los modelos

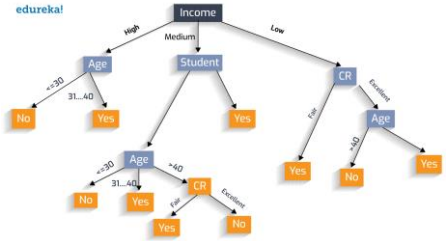
Predicción del rendimiento académico

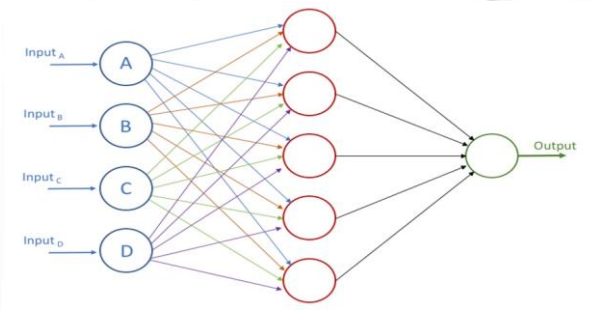
Metodologías empleadas

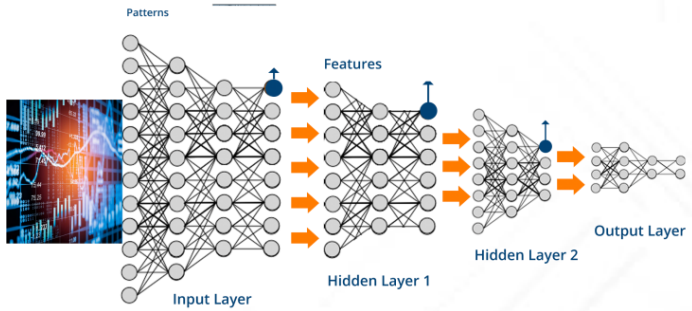
- Clasificación (cualitativa) vs regresión (numéricas).
- Modelos caja blanca (interpretables) vs caja negra (precisos)
- Modelos clásicos (estadística tradicional, *machine learning*) vs actuales (*deep learning*)

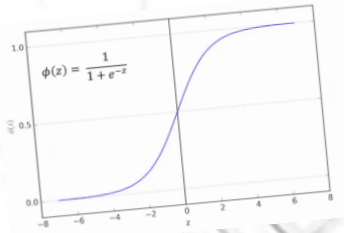






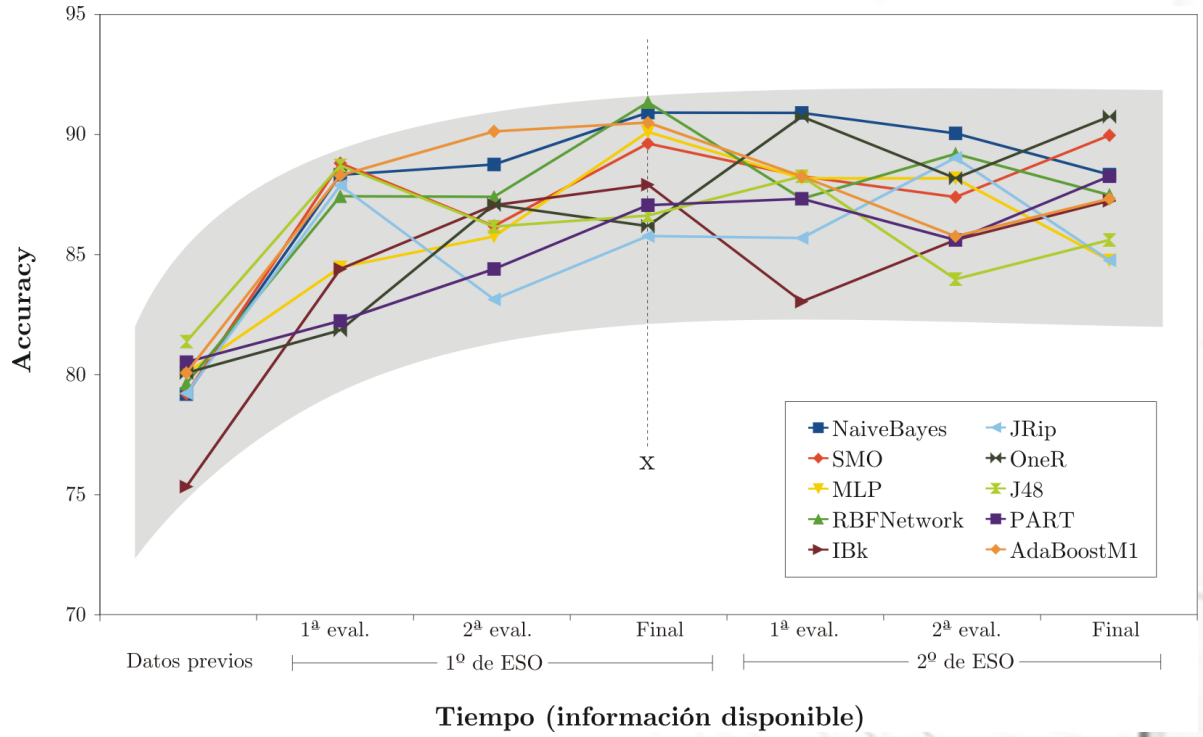






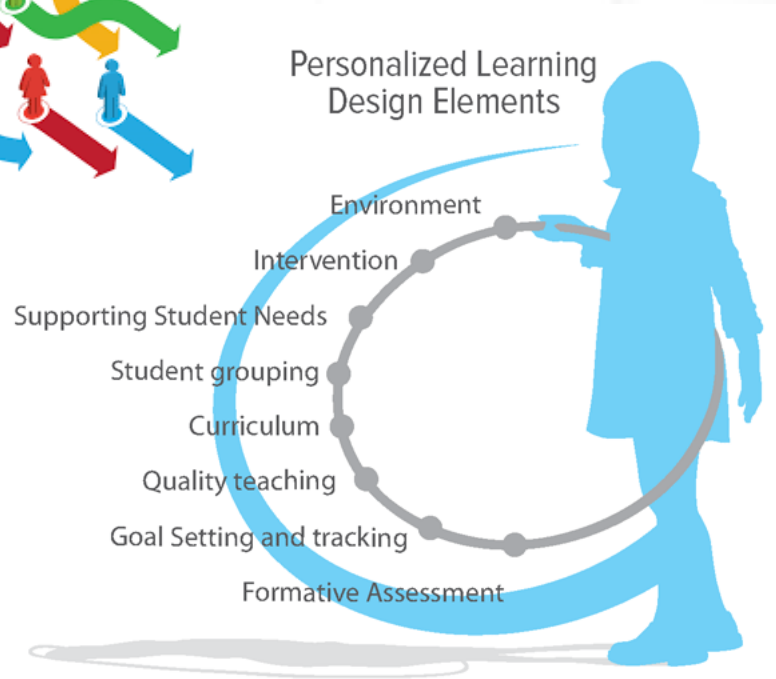
Predicción del rendimiento académico
Retos abiertos

- Ventana temporal
- Datos procedentes de distintas fuentes
  - Datos tabulares vs. relacionales
  - Representaciones de datos flexibles
- Validez de los modelos
  - Caducidad de los modelos?
  - Big Data vs. Small Data



Personalización de la enseñanza
Concepto

- Término muy amplio:
  - Personalización de actividades
  - Personalización de contenidos
  - Personalización de caminos de aprendizaje
- Objetivos:
  - Mejorar la experiencia de los estudiantes en el sistema
  - Mejorar la eficiencia del proceso de enseñanza-aprendizaje:
    - Mejora del rendimiento
    - Metas perseguidas



Personalización de la enseñanza
Metodologías empleadas

Soluciones basadas en los principios de los Sistemas de Recomendación:

- Filtrado colaborativo:
  - Agrupamiento (*clustering*) basado en características
  - Análisis de redes sociales
- Modelos basados en contenido:
  - Modelos de clasificación convencionales
  - Análisis de sentimientos
- Sistemas híbridos

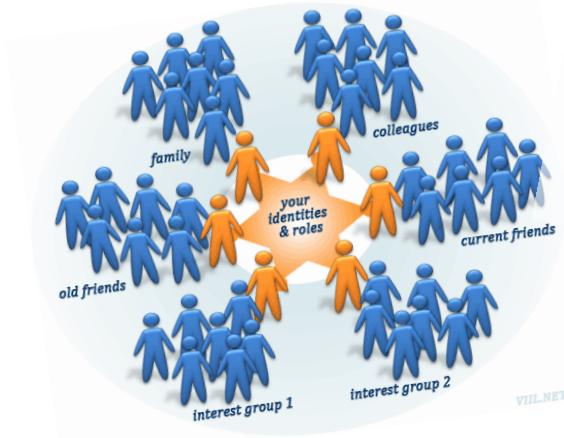
Personalización de la enseñanza

Recomendación de itinerarios / asignaturas en alumnos universitarios

- Problema reciente
- Doble objetivo:
  - Obtener un modelo que permita recomendar a los estudiantes un **itinerario académico** (qué asignaturas optativas elegir y en qué orden elegirlas) en función de determinados objetivos: calificación media, especialización, esfuerzo realizado…
  - Establecer **índices de dificultad de asignaturas** basados en la experiencia previa. Estos índices pueden ser utilizados en la construcción del modelo indicado anteriormente
- Información disponible:
  - Registros académicos de una Universidad durante los 10 últimos años. El número de registros es muy grande (más de 1000000 en este período de tiempo).

Formación de grupos / equipos de trabajo

- Objetivos: Identificar alumnos con características similares, para personalizar el aprendizaje o identificar alumnos compatibles y complementarios para formar equipos de trabajo
- Las técnicas de agrupamiento (*clustering*) permiten encontrar individuos con características similares
- También las técnicas de análisis de redes sociales pueden usarse para esta finalidad, proporcionando medidas de la reputación de los individuos, afinidad con sus iguales, detección de grupos, …

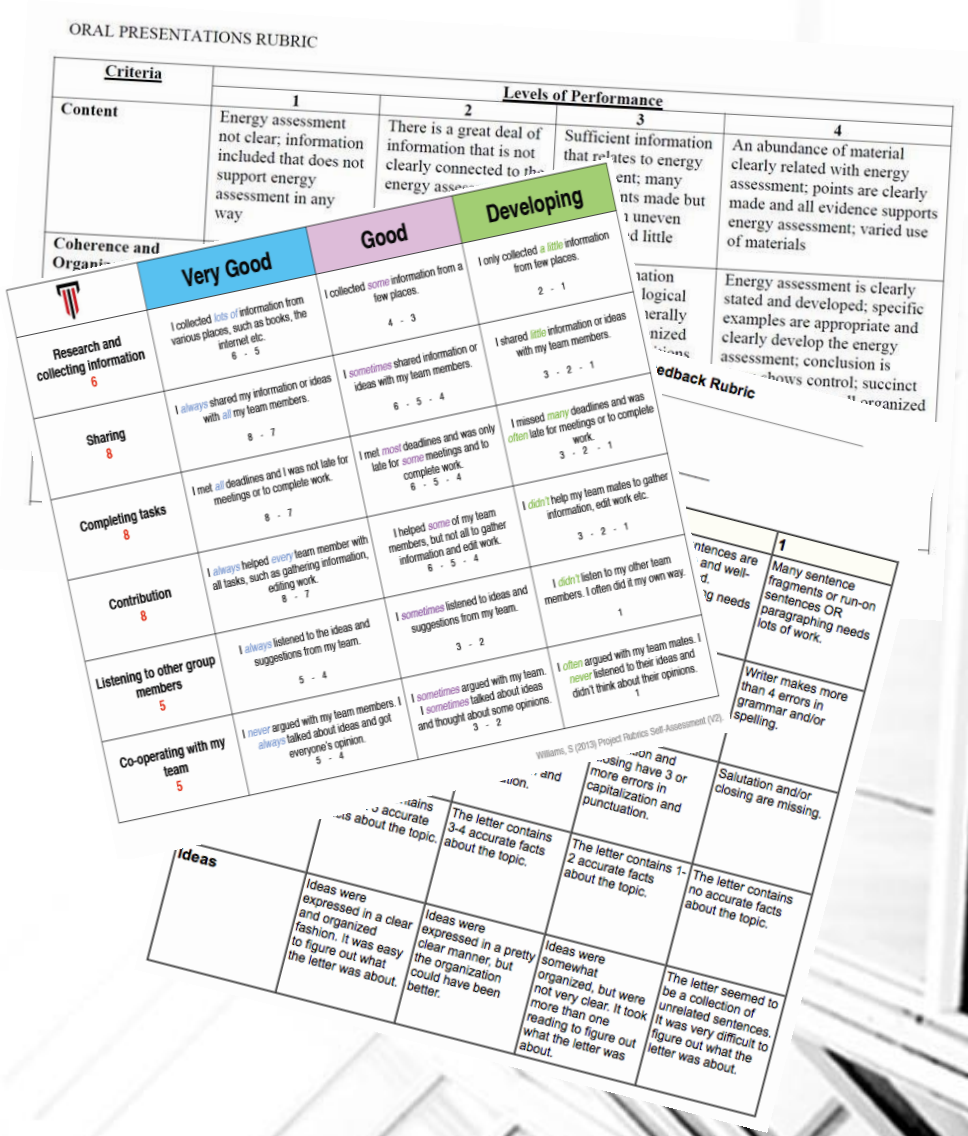


Modelado de la evaluación por pares y de la autoevaluación

- La aparición de los MOOCs (*Massive Open Online Courses*) plantea un gran número de retos asociados al número de estudiantes.
- La evaluación automática es un gran problema en algunas materias:
  - Ensayos,
  - Traducciones
  - Programas de ordenador
  - ...
- Nuevos modelos de evaluación:
  - Evaluación por pares
  - Autoevaluación

Modelado de la evaluación por pares y de la autoevaluación

- Tanto en los sistemas de evaluación por pares como en los de autoevaluación es frecuente el uso de rúbricas (plantillas de evaluación).
- En ambos sistemas hay desviaciones entre las evaluaciones de los compañeros o del propio alumno y las evaluaciones del profesor.
- Se establecen modelos matemáticos que modelan dichas desviaciones:
  - Definición de una puntuación que mida la “reputación” de cada evaluador.



Nuevos Retos

Problemas aún activos en EDS (EDM/LA)

- Desarrollo de herramientas específicas para entornos educativos y profesionales de la educación
- Datos dinámicos: minería de procesos y flujos de datos
- Análisis de video y sensores
- Interés del estudiante: emoción, afecto y elección
- Integración de las técnicas de ciencia de datos con la teoría educativa
- Mejorando el soporte al profesorado: efectividad de los materiales, metodología empleada, ...
- Análisis de datos no estructurados: documentos, foros, chats, sesiones de tutoría...
- Análisis de las redes sociales: Twitter, Facebook,...

Nuevos desafíos

Ryan Baker propone en un trabajo reciente, los “Premios Baker de Analítica de Aprendizaje”. Se trata de buscar soluciones a los siguientes problemas:

- Evaluación de los modelos desarrollados.
- Desarrollo de modelos interpretables.
- Aplicabilidad de los modelos
- Transferibilidad de los modelos
- Generalidad de los modelos

Conclusiones

Conclusiones

- La Ciencia de Datos puede utilizarse con éxito a la mejora de los sistemas educativos
- Se trata de una disciplina ya consolidada, pero que progresa muy deprisa, planteando problemas cada vez más complejos, de mayor interés
- La aparición de datos masivos (*big data*) en los sistemas educativos, y la incorporación de tecnologías emergentes (realidad virtual y aumentada, internet de las cosas, gemelos digitales, etc.) están dando una importancia creciente a la formación permanente.
- Las tecnologías de enseñanza online, soportadas por la inteligencia artificial y la computación cognitiva deben jugar un papel clave dentro de este nuevo escenario.

Ciencia de Datos en Educación

Sebastián Ventura Soto

Grupo de Investigación *Knowledge Discovery and Intelligent Systems*

Universidad de Córdoba