Asignatura: Análisis Exploratorio de Datos

Presentación de la Asignatura y Revisión del Programa

Profesor

- José María Matías, jmmatias@uvigo.es.
  - Despacho nº 8, ETSEM, campus de Vigo, Lagoas, 36310, Vigo.
  - Tfno. 986 812190.

Ficha del alumno

- Complimentar Ficha del Alumno. Se enviará el formulario a remitir por mail a la dirección jmmatias@uvigo.es.

Desarrollo de la Docencia

- Asignatura orientada a Tareas: todo lo que se estudie tendrá como finalidad la realización de distintos trabajos de prácticas.
  - Periódicamente, cada semana o dos, se planteará un boletín de tareas orientadas a la resolución de problemas prácticos: lectura, instalación de software, tratamiento específico de datos, etc.
  - Cada boletín será autosuficiente de tal forma que pueda realizarse a distancia.
  - Al margen de ello, en clase, el primer día del período en cuestión se presenta el boletín y se aportan ideas generales sobre qué se pretende y cómo se realizan las tareas.
  - El último día del período correspondiente se plantean dudas y se realizan comentarios colectivamente.
- Por tanto, las clases tendrán el carácter de tutorías. Además, se podrán hacer consultas por mail siempre que sean operativas (difícil si se refieren a uso de software) con los siguientes horarios de respuesta:
  - Lunes: de 10:00h a 13:00h.
  - Miércoles: de 11:00h a 14:00h.
- Habrá un examen final presencial con apuntes y libros reproduciendo el ambiente de los trabajos. No está claro todavía si será tipo test y si será en un aula con ordenadores para realizarlo en máquina con software instalado.
- El cuatrimestre consiste en 13 semanas aproximadamente. La docencia se desarrollará de acuerdo con las fases de Descubrimiento (Construcción) de Conocimiento (Knowledge Discovery) de las cuáles las más importantes para nosotros son las fases de Exploración, Selección, Limpieza, Transformación, Minería de Datos y Evaluación.

Alcance de la Asignatura

- No se limita a Estadística Descriptiva.
- No se limita a la fase de Exploración de un proyecto de Minería de Datos sino que abarcará las fases de contenido estadístico de un proyecto de este tipo.

- Estadística Descriptiva → Análisis Exploratorio de Datos → Minería de Datos con una orientación final encuadrada en el problema estadístico de aprendizaje.

Bibliografía más relevante

- Básico para seguir (Castellano): Hernández Orallo et al., Introducción a la minería de datos, Pearson - Prentice Hall. Última edición, 2007 (valen otras). Hay en Biblioteca de Vigo.
- Complementará al anterior (castellano): C. Pérez López. Minería de Datos. Thomson (mejor en la visión general, detalle/formalismo, pero no en amplitud y además revisa software comercial inaccesible).
- Referencia (Inglés): Han y Kamber. Data Mining: Concept and Techniques. Morgan Kaufman. 2001. No está en el programa pero de los mejores.
- Sencillo de carga matemática y en Inglés: Larose. Discovering Knowledge in Data. An Introduction to Data Mining. Wiley, 2005.
- Referencia (inglés) sobre técnicas con consciencia estadística. Sirve de biblioteca de técnicas: Hastie, Friedman, Tibshirani. The Elements of Statistical Learning. Springer, 2001.
- Clásico por la relevancia del software libre asociado (Weka): Witten I. H. y Frank E. (2005) Data Mining. Practical Machine Learning Tools and Techniques. Elsevier.
- Ver en Internet libros con títulos que incluyan “Data Mining”, “Machine Learning”, “Pattern Recognition”, “Exploratory Data Analysis”. Relevantes también todos los de Análisis Multivariante.
- Libros sobre Statistical Learning Theory: poseen un enfoque más rico y potente pero su lectura más allá del capítulo segundo requiere más voluntad y formación matemática previa.
  - Vapnik. The Nature of the Statistical Learning Theory. Springer, 2000.
  - Cherkassky y Mulier. Learning from Data. Wiley, 1998.

Software

- Utilizaremos el que mejor se adecúe a cada tarea entre los disponibles con licencia de campus en alguna universidad de Galicia y entre los gratuitos disponibles en Internet.