

Guía docente de Asignatura – Máster en minería de datos e inteligencia de negocios

Datos generales de la asignatura

Asignatura:	Técnicas y Metodología de la minería de datos (SEMMA) - 606538
Curso académico:	2023-24
Carácter:	Obligatoria
Curso:	Primero
Semestre:	1
Créditos ECTS	
Presenciales:	1,8
No presenciales:	4,2
Total	6,0
Actividades docentes	
Clases teóricas:	50%
Clases prácticas:	50%
Total:	100%
Departamentos responsables:	Estadística y Ciencia de los Datos
Profesores:	Cesar Perez Lopez
Profesor Coordinador:	Cesar Perez Lopez

Datos específicos de la asignatura

Breve descriptor:	Esta asignatura se contempla como una introducción al análisis de datos integrado utilizando software estadístico amigable (Clementine, Miner SAS). Se describen técnicas estadísticas habituales con el objetivo de que se conozca el entorno informático. Se hace especial hincapié en la presentación y resolución de problemas prácticos a los que se les va aplicando diferentes técnicas, entendiéndose como un procedimiento sistemático de trabajo. Exploración y depuración de datos, resumen de la información y predicción de resultados. El método de trabajo consistirá en utilizar dos grandes bases de datos y trabajar con ellas de forma integrada. Será necesario repetir más de una vez los diferentes pasos que configuran la filosofía SEMMA con el objeto de que al final nos quedemos satisfechos con los resultados de los análisis, confirmando se cumplan todas las hipótesis asociadas a los modelos y técnicas empleadas. Además, se buscará que el alumno se ayude a través de diagramas de flujo que describan adecuadamente el proceso del tratamiento de los datos. Por esta razón utilizaremos el interfaz gráfico presente en el software comercial más generalizado. Las técnicas a las que se dará prioridad en esta asignatura son los métodos de selección de variables, la regresión lineal, la regresión logística y los árboles de regresión.
Requisitos:	
Competencias	
Generales:	El alumno debe saber depurar la información, conocer, aplicar, comprobar e interpretar los resultados que se deriven de los análisis realizados. Debe dominar el software de aplicación de la metodología SEMMA. Realizar diagramas informativos que expliquen la secuencia de técnicas a utilizar.
Transversales:	
Específicas:	
Objetivos	
Contenidos	1) Metodologías de Minería de Datos y Machine Learning: SEMMA

- 2) Introducción al Enterprise Miner de SAS.
- 3) Técnicas de exploración y depuración de datos en SAS Miner.
- 4) Técnicas de selección de variables y visualización. Análisis no supervisado.
- 5) Técnicas de modelización relativas a clasificación supervisada (binaria): k-vecinos (kNN), árboles de clasificación, regresión logística
- 6) Técnicas de modelización relativas a predicción continua: regresión lineal, árboles de regresión.

Evaluación

El alumnado será evaluado continuamente a lo largo del curso. Será obligatoria la entrega a lo largo del curso de prácticas individuales cuya resolución se llevará a cabo dentro del aula, y cuyo peso global supondrá el 40% de la nota final. Además, se plantearán entregas grupales cuyo peso en la nota final será del 40%, y cuya resolución se llevará a cabo tanto dentro como fuera del aula. El 20% restante se asignará en función de la nota obtenida en una última práctica grupal final cuya resolución se llevará a cabo dentro del aula, y que abarcará los contenidos de toda la asignatura.

El alumnado deberá entregar obligatoriamente todas las entregas individuales y participar en, al menos, una de las entregas grupales planteadas. En caso de que el alumno no supere el 3 sobre 10 en las entregas individuales, y/o la nota media de las entregas individuales sea un 50% inferior a la nota obtenida en las entregas grupales, se le podrá requerir la exposición de contenidos en una tutoría individual que será evaluada. En caso de que un alumno no haya superado la asignatura por este método, haya faltado a más de un tercio de las clases, o quiera mejorar su calificación se podrá presentar a un examen final y su nota será el 100% de la nota del curso.

Bibliografía

- Hastie, Trevor; Tibshirani, Robert y Friedman, Jerome (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.
- Kuhn, Max y Johnson, Kjell (2016). Applied Predictive Modelling. Springer.
- Gareth, James et. al (2015). An introduction to statistical learning : with applications in R. Springer.
- Álvarez-Liébana, J. (2021). Manual introductorio online. «Aprendiendo R sin morir en el intento». Disponible gratuitamente en <https://datosdelaplace.github.io/courses-intro-R/>
- Álvarez-Liébana, J. (2021). Curso de visualización de datos. «Analizando información, visualizando datos, contando historias». Disponible gratuitamente en <https://datosdelaplace.github.io/courses-ECI-2022/>
- Swirl stats: aprendizaje básico de R dentro del propio R. Disponible gratuitamente en <https://swirlstats.com/>
- Wright, C., Ellis, S.E., Hicks, S. C., Peng, R. D. (2021). «Tidyverse skills for data science». Disponible gratuitamente en <https://jhudatascience.org/tidyversecourse/>
- Matloff, N (2011). The Art of R Programming. A tour of statistical software design. Disponible como recurso electrónico en la biblioteca de la UCM.
- Crawley, M.J. The R book. Disponible como recurso electrónico en la biblioteca de la UCM
- Santin González, Daniel y Pérez, López, César (2007). Minería de datos, técnicas y herramientas. Tomson.
- Cerrito B. Patricia (2006). Introduction to Data Mining Using SAS Enterprise Miner. SAS Institute.