

Estadística avanzada y Minería de datos / Advanced Statistics and Data Mining

Número total de créditos ECTS	6	
Tipología	Obligatoria	
Organización temporal	1er. curso/ 1er. semestre	
Modalidad	Presencial	
Idioma	Castellano	
Contenidos	<ul style="list-style-type: none"> · Técnicas de descripción y modelado de datos complejos: muestreo, training, test set. · Modelos de Regresión: lineal, no lineal, logística. Regularización: ridge y lasso. · Métodos núcleo y splines. Función de pérdida. · Aprendizaje Bayesiano: análisis bayesiano, inferencia bayesiana, métodos MCMC, modelado bayesiano e inferencia, modelos jerárquicos Bayes. · Modelos gráficos probabilistas. Bayesianos, cadenas de Markov, Filtros de Kalman, redes de creencia. · Modelado de funciones de densidad de probabilidad · Series temporales: introducción, descomposición, medias móviles, ARIMA, estacionariedad, predicción. · Optimización para grandes volúmenes de datos: Programación lineal. Programación cuadrática. Programación no lineal. Heurísticas. Metaheurísticas. · Proyectos de minería de datos · Proceso de KDD · Técnicas de pre-procesamiento de datos · Métodos de clasificación · Sistemas de recomendación · Análisis de datos con cubos y modelos de minería · Evaluación y selección de modelos: matriz de confusión, métricas, costes. Curvas ROC <p>-----</p> <ul style="list-style-type: none"> · Techniques for describing and modelling complex data: sampling, training, test set. · Regression models: linear, non-linear, logistic. Regularisation: ridge and lasso. · Kernel and spline methods. Loss function. · Bayesian learning: Bayesian analysis, Bayesian inference, MCMC methods, Bayesian modelling and inference, hierarchical Bayes models. · Probabilistic graphical models. Bayesian, Markov chains, Kalman filters, belief networks. · Probability density function modelling · Time series: introduction, decomposition, moving averages, ARIMA, stationarity, forecasting. · Optimisation for large data volumes: Linear programming. Quadratic programming. Non-linear programming. Heuristics. Metaheuristics. · Data mining projects · KDD process · Data pre-processing techniques · Classification methods · Recommender systems · Data analysis with mining cubes and models · Model evaluation and selection: confusion matrix, metrics, costs. ROC curves. 	
Resultados de aprendizaje TÍTULO	Conocimientos y contenidos	CC1 Comprender las técnicas y herramientas de programación especializada en modelado, gestión, procesamiento y análisis de datos en entornos de Big Data. //Having a knowledge of Understand specialized of programming techniques and tools

		specialized specialized in for modelling, management, processing and analysis of data in Big Data environments.
	Habilidades y destrezas	HD4 Utilizar de forma integrada conjuntamente métodos estadísticos avanzados, recursos informáticos y de diagnóstico estratégico para elaborar soluciones Big Data de alto valor añadido. //Jointly use advanced statistical methods, IT and strategic diagnostic resources to develop high value-added Big Data solutions. HD5 Sintetizar los resultados de los modelos de análisis de datos en un contexto de toma de decisiones. //Synthesise the results of data analysis models in a decision-making context.
	Competencias	CP2 Aplicar los principales métodos de minería de datos para extraer características y contextualizar información con el fin de facilitar la posterior toma de decisiones. //Apply the main data mining methods extract features and to contextualise information in order to facilitate subsequent decision making. CP3 Aplicar técnicas de generación de visualizaciones de datos para su análisis y comunicación de resultados orientados a la toma de decisiones. //Apply data visualisations techniques for data analysis and communication of results aimed at decision-making. CP6 Identificar la solución más óptima para un caso propuesto que utilice modelos y técnicas de Datos Masivos, en términos de eficiencia, eficacia e interpretación de resultados. //Identify the Most optimal solution for a Big Data issue, in terms of efficiency, effectiveness and interpretation of results.

Resultados de aprendizaje ASIGNATURA

- RA.1.- Conocer los principios básicos de las técnicas de diseño en ciencia de datos.
- RA.2.- Identificar los descriptores más habituales para un conjunto de datos y valorar su aplicabilidad para una base de datos conocida.
- RA.3.- Conocer las distintas técnicas de Regresión: lineal, no lineal y logística. Entender sus aplicaciones en ciencia de datos: predicción o clasificación de clases.
- RA.4.- Conocer las técnicas de extracción de características para reducir la dimensionalidad de un conjunto de datos y mejorar la eficacia de los algoritmos de minería de datos.
- RA.5.- Implantar una estrategia integrada de técnicas estadísticas para el desarrollo de sistemas descriptivos y predictivos.
- RA.6.- Aplicar las herramientas informáticas para el análisis estadístico y el proceso masivo de datos.
- RA.7.- Preparar datos mediante técnicas de preprocesamiento para su posterior procesamiento.
- RA.8.- Entender y aplicar las principales técnicas de clasificación, agrupamiento de datos y evaluación de modelos.
- RA.9.- Aplicar técnicas estadísticas y de minería de datos con el fin de generar alternativas de decisión a través de patrones de comportamiento identificados.
- RA.10.- Diseñar y ejecutar una visualización, desde la recolección de datos hasta su final representación gráfica.
- RA.11.- Adecuar una visualización concreta a la finalidad de la misma y su público objetivo.
- RA.12.- Seleccionar la herramienta más adecuada para la ejecución de una visualización efectiva.

-
- RA.1.- Know the basic principles of design techniques in data science.
 - RA.2.- Identify the most common descriptors for a dataset and assess their applicability to a known database.
 - RA.3.- Know the different techniques of regression: linear, non-linear and logistic. To understand their applications in data science: prediction or class classification.
 - RA.4.- Know the feature extraction techniques to reduce the dimensionality of a dataset and improve the efficiency of data mining algorithms.
 - RA.5.- Implement an integrated strategy of statistical techniques for the development of descriptive and predictive systems.

RA.6.- Apply computer tools for statistical analysis and massive data processing.
 RA.7.- Prepare data by means of pre-processing techniques for their subsequent processing.
 RA.8.- Understand and apply the main classification techniques, data grouping and model evaluation.
 RA.9.- Apply statistical and data mining techniques in order to generate decision alternatives through identified behavioural patterns.
 RA.10.- Design and execute a visualisation, from data collection to its final graphical representation.
 RA.11.- Adapt a specific visualisation to its purpose and target audience. RA.12.- Select the most appropriate tool for the execution of an effective visualisation.

Actividades formativas	Horas totales
AF1. Lección magistral participativa/ Participative master class	20
AF2. Prácticas de laboratorio/ Laboratory practicals	14
AF3. Resolución de problemas/ Problem solving	14
AF8. Trabajo autónomo/Autonomous work	98
AF7. Evaluación/Evaluation	4
Total	150

* Sistemas de evaluación	MÍNIMO	MÁXIMO
La Evaluación del Portafolio. Informe sobre resolución de problemas o estudio de casos / Portfolio assessment. Report on problem solving or case studies	20	30
Prueba final / Final test.	40	60
Participación en grupos de debate /Participation in discussion groups	10	10
Evaluación del portafolio. Informe sobre Diseño de Proyectos / Portfolio evaluation. Report on Project Design	10	20

* El peso de cada sistema de evaluación quedará definido en la guía docente de la asignatura.