



GRADO

# GUÍA DE ESTUDIO DE LA ASIGNATURA

## MINERÍA DE DATOS



2017-2018

GRADO EN INGENIERÍA EN TECNOLOGÍAS DE LA INFORMACIÓN

UNIVERSIDAD NACIONAL DE EDUCACIÓN A DISTANCIA

## 1.- PLAN DE TRABAJO

El plan de trabajo de la asignatura aparece descrito en la Tabla 1. En él se puede ver la secuencia de tareas que el estudiante tendrá que realizar a lo largo del curso.

El contenido de la asignatura ha sido diseñado no con el objetivo de dar un panorama completo del área, sino con el de proporcionar al alumno los conceptos clave que subyacen a todos los desarrollos del campo. Sería imposible revisar todas las técnicas desarrolladas durante las últimas décadas con el objetivo de automatizar el reconocimiento de patrones o el descubrimiento de conocimiento en bases de datos: modelos gráficos probabilistas, modelos basados en reglas, redes neuronales, máquinas de vectores soporte, procesos gaussianos modelos basados en árboles de decisión/regresión y un sinnúmero de alternativas. De hacerlo, es seguro que el nivel de comprensión de cada uno de los métodos sería cuando menos superficial.

El equipo docente ha elegido por el contrario unos contenidos básicos (que no simples) que permitan al alumno en el futuro adquirir conocimientos avanzados de cualquier especialidad del área de la Minería de Datos con suma facilidad. Lo más difícil estará hecho: se tratará únicamente de aplicar los conceptos y destrezas adquiridos a una metodología particular. Asimismo, los alumnos interesados en continuar su formación de postgrado en el área de la Minería de Datos estarán igualmente preparados para profundizar en los diferentes aspectos del campo de investigación.

El diseño de la guía de estudio alterna bloques de lectura con bloques de ejercicios. Finalmente, el equipo docente ofrece la posibilidad de realizar una práctica voluntaria y evaluable basada en la experimentación con el programa estadístico [R](#). La experimentación sugerida está diseñada para consolidar los conocimientos adquiridos desde una perspectiva aplicada.

La mayor parte de las tareas consisten en la lectura de los apartados (epígrafes) del libro de texto básico de la asignatura, y en la realización de unas actividades básicas no evaluables tras la lectura de cada epígrafe, que le ayudarán a asimilar mejor los contenidos. Las actividades básicas no evaluables incluyen aquellas que aparecen al final de cada epígrafe del libro y pueden ser de auto-evaluación (ejercicios resueltos) o de descubrimiento (ejercicios no resueltos).

El equipo docente propondrá cada año dos subconjuntos de ejercicios tomados del texto base. El primero estará compuesto de ejercicios cuya respuesta se encuentra disponible en el sitio web del libro (auto-evaluación); el segundo contendrá ejercicios cuya respuesta no está disponible a través de Internet y de entre los que se seleccionarán los enunciados del examen (descubrimiento). Nosotros recomendamos que los estudiantes intenten resolver tanto los unos como los otros, de la misma manera: a ciegas, sin mirar la respuesta (cuando ésta esté disponible) hasta después de haber resuelto el enunciado y sólo como comprobación.

El 10% de la nota final de la asignatura corresponderá a la evaluación de por parte de los tutores de un conjunto de 3 ejercicios de descubrimiento elegidos por el estudiante de entre los propuestos por el equipo docente.

Un último tipo de tarea incluye la realización de una práctica de experimentación numérica evaluable. Dicha práctica será evaluada por los tutores, y podrá suponer hasta 2 puntos sobre 10 en la nota final. La nota final se calculará sumando a la nota del examen la puntuación de la práctica siempre y cuando ésta última supere los 4 puntos sobre 10. Si la suma de ambas notas supera los 10 puntos, la nota evidentemente será de 10. Su objetivo (el de la práctica evaluable) es facilitar que el alumno adquiera familiaridad con los casos prácticos de experimentación numérica a los que se les aplica todo el bagaje conceptual adquirido durante el curso. El enunciado de la práctica se hará público cada año a comienzo de curso.

Es importante que el alumno sepa que las actividades evaluables sólo se realizarán y evaluarán en el cuatrimestre en el que se imparte la asignatura y que, por tanto, en la convocatoria de Septiembre, se mantendrá la nota obtenida en Febrero en dichas actividades.

No le proporcionamos un calendario de trabajo por semanas, pues creemos que el alumno puede construirse fácilmente a partir de la tabla presentada y que, además, dependerá de cuándo el alumno comience a estudiar la asignatura. No obstante, consideramos que la mejor forma de cumplir con la planificación marcada en la tabla es comenzar a estudiar desde el comienzo de curso, pues creemos que la

asignatura es muy fácil si se lleva paso a paso, pero muy difícil si se pretende preparar en pocas semanas antes del examen.

La tabla se ajusta al número de créditos ECTS que corresponden a la asignatura, de forma que para 6 créditos el número de horas ha de ser  $25 \times 6 = 150$  horas.

En cualquier caso, la realimentación que el equipo docente obtenga de los alumnos a lo largo del curso nos servirá para dar una planificación lo más realista posible que consiga los mejores resultados de aprendizaje en cursos futuros.

## **ORIENTACIONES PARA LA REALIZACIÓN DEL PLAN DE ACTIVIDADES**

El equipo docente recomienda la lectura completa de cada tema antes de empezar a abordar tanto los ejercicios de auto-evaluación como los de descubrimiento. Sólo en el caso del primer bloque nos encontramos una correspondencia no unívoca con un tema del texto base. Este primer bloque, en realidad, se corresponde con los temas 1 y 2 de dicho texto, por lo que se pueden intercalar los ejercicios correspondientes a los sub-bloques 1.1-1.6 antes de empezar a estudiar el 1.7.

Como se indica en el apartado 1 de esta parte II de la guía de la asignatura, los enunciados del examen se corresponderán con algunos de los ejercicios de descubrimiento propuestos a comienzo de curso por el equipo docente. Por lo tanto, no habrá preguntas que exijan la memorización de demostraciones. Si es necesario, el equipo docente proporcionará definiciones complejas que sean necesarias para la resolución de los enunciados de examen. Consideramos que lo más importante no es la memorización de fórmulas sino la comprensión de los conceptos clave del área y la adquisición de las destrezas algebraicas necesarias para manipular fórmulas.

Finalmente, el equipo docente permitirá que los estudiantes utilicen como material auxiliar el texto base de la asignatura sin anotaciones de ningún tipo.

Recomendamos pues que los estudiantes aborden la resolución de todos los ejercicios de descubrimiento propuestos por el equipo docente como la vía más segura de preparación para el examen.

Para la práctica voluntaria, los criterios de evaluación se recogen en la siguiente tabla:

Indicador	Nivel de ejecución			Totales parciales
Claridad y pulcritud en la presentación de los resultados	Texto en correcto castellano y sin faltas de ortografía	Lo anterior + <u>claridad en la exposición</u> de los objetivos, de la metodología empleada y de los resultados obtenidos.	Todo lo anterior + división clara en secciones, con gráficas ilustrativas y claras, y con las tablas <b>necesarias (sin exagerar)</b> para clarificar el texto.	10,00%
Adecuación del código a las normas de Ingeniería del Software	Código comentado, bien estructurado, y limpio de código basura.	Lo anterior + parámetros de ejecución leídos de fichero de configuración. Sin hard-coding. Eficiente en términos de CPU y uso de memoria.	Todo lo anterior + código reutilizable, modular y con tests unitarios.	15,00%
Obtención de los objetivos propuestos	La práctica no está completa o, en caso de estarlo, presenta un número reducido de secciones que contienen errores graves,	La práctica está bien resuelta salvo fallos de menor importancia	La práctica proporciona todos los objetivos propuestos sin errores o fallos de ejecución.	70,00%
Originalidad de la solución	No existe	No existe	Solución muy optimizada o que se aparta de la solución más obvia.	5,00%

Para los enunciados de la prueba presencial:

Indicador	Nivel de ejecución			Totales parciales
Claridad y pulcritud en la presentación de los resultados	Texto en correcto castellano y sin faltas de ortografía. Respuesta legible.	Lo anterior + <u>claridad en la exposición</u> de los objetivos, de la metodología empleada y de los resultados obtenidos.	Todo lo anterior + división clara en secciones, con gráficas ilustrativas y claras, y con las tablas <b>necesarias (sin exagerar)</b> para clarificar el texto.	30,00%
Obtención de los objetivos propuestos	La respuesta es incompleta, contiene errores graves en algún punto pero que no invalidan	La respuesta es correcta (obtiene los objetivos perseguidos) pero lo hace de manera poco rigurosa, sin	La respuesta consigue los objetivos de manera clara y rigurosa, la solución está claramente explicada	70,00%

	desarrollo previo. aclara los pasos de la solución, cometiendo errores u omisiones de poca importancia que se pretende evaluar.	y no contiene errores de ningún tipo.	
--	---	---------------------------------------	--