26-27

GRADO EN INGENIERÍA EN INTELIGENCIA ARTIFICIAL
SEGUNDO CURSO

GUÍA DE ESTUDIO PÚBLICA

FUNDAMENTOS DE MODELADO ESTADÍSTICO DE DATOS

CÓDIGO 71032021

UNED

26-27

FUNDAMENTOS DE MODELADO ESTADÍSTICO DE DATOS

CÓDIGO 71032021

ÍNDICE

PRESENTACIÓN Y CONTEXTUALIZACIÓN
REQUISITOS Y/O RECOMENDACIONES PARA CURSAR LA ASIGNATURA
EQUIPO DOCENTE
HORARIO DE ATENCIÓN AL ESTUDIANTE
TUTORIZACIÓN EN CENTROS ASOCIADOS
COMPETENCIAS QUE ADQUIERE EL ESTUDIANTE
RESULTADOS DE APRENDIZAJE
CONTENIDOS
METODOLOGÍA
SISTEMA DE EVALUACIÓN
BIBLIOGRAFÍA BÁSICA
BIBLIOGRAFÍA COMPLEMENTARIA
RECURSOS DE APOYO Y WEBGRAFÍA
IGUALDAD DE GÉNERO

NOMBRE DE LA ASIGNATURA	FUNDAMENTOS DE MODELADO ESTADÍSTICO DE DATOS
CÓDIGO	71032021
CURSO ACADÉMICO	2026/2027
DEPARTAMENTO	INTELIGENCIA ARTIFICIAL
TÍTULO EN QUE SE IMPARTE	GRADO EN INGENIERÍA EN INTELIGENCIA ARTIFICIAL
CURSO	SEGUNDO CURSO
PERIODO	SEMESTRE 1
Nº ETCS	6
HORAS	150.0
IDIOMAS EN QUE SE IMPARTE	CASTELLANO

PRESENTACIÓN Y CONTEXTUALIZACIÓN

La asignatura "Fundamentos de modelado estadístico de datos" (FMED) se enmarca dentro del aprendizaje supervisado, donde se modeliza una variable respuesta (variable resultado o dependiente) en función de ciertas variables explicativas (predictores o variables independientes). Los principales modelos que se estudian son los modelos de regresión y los modelos de clasificación.

La asignatura pretende hacer asequible los fundamentos teóricos del funcionamiento interno de los modelos, sin olvidar la parte práctica de su utilización mediante software estadístico libre.

La asignatura FMED forma parte del Grado en Ingeniería en Inteligencia Artificial (GIIA), consta de 6 ECTS, es de carácter obligatorio y pertenece a la materia de "Matemáticas para la IA", que está constituida por las asignaturas "Fundamentos de cálculo para la IA", "Fundamentos algebraicos para la IA", "Lógica y estructuras discretas", "Fundamentos de Estadística", "Fundamentos de modelado estadístico de datos", "Métodos analíticos para la toma de decisiones" y "Ampliación de métodos probabilistas".

FMED contribuye al futuro perfil profesional y/o investigador del estudiante en el aspecto fundamental del análisis de los datos y la creación de modelos predictivos. FMED profundiza en los aspectos tratados en la asignatura de "Fundamentos de estadística para la IA" y proporciona los fundamentos estadísticos teóricos básicos necesarios para las asignaturas de "Aprendizaje Automático I", "Aprendizaje Automático II" y "Aprendizaje automático III".

La presente guía contiene información de carácter general sobre la asignatura: requisitos y recomendaciones, equipo docente, horario de atención, resultados de aprendizaje, contenidos, metodología, plan de trabajo, sistema de evaluación, bibliografía básica, bibliografía complementaria, recursos de apoyo y glosario.

Es de destacar que en esta asignatura, la interacción con los alumnos es constante en la plataforma virtual y que se envían noticias sobre dicha interacción al correo de la UNED del alumno, por lo que conviene acceder a la plataforma y a dicho correo de forma regular (al

menos una vez por semana). Se ha comprobado en otras asignaturas que esta forma de interacción con los alumnos es muy bien valorada en las encuestas anónimas que realizan los estudiantes (en esta asignatura, al ser el pimer año que se oferta no hay todavía datos).

REQUISITOS Y/O RECOMENDACIONES PARA CURSAR LA ASIGNATURA

Los requisitos inciales son haber superado previamente las asignaturas:

•"Fundamentos de cálculo para la IA".

•"Fundamentos algebraicos para la IA".

•"Fundamentos de estadística para la IA".

No obstante, se repasarán conceptos básicos estadísticos en los primeros temas y se darán ayudas sobre aspectos matemáticos cuando se necesite. Por otra parte, es recomendable conocer (leer y escribir) el inglés técnico.

Como recomendación general para esta asignatura conviene tener en cuenta lo siguiente:

•El foro de la plataforma virtual es un espacio vivo en el que el equipo docente está presente de forma continua para ayudar a los estudiantes a progresar en su estudio.

•En el foro de la plataforma virtual, se va marcando un ritmo de estudio semanal para poder llevar al día la asignatura (para aquéllos que puedan y/o deseen).

•En todo momento es factible reengancharse a la asignatura gracias a los resúmenes que se van poniendo en la plataforma virtual dentro de las preguntas más frecuentes (P+F).

•**Es muy importante asistir en directo o en diferido a la sesión telemática de bienvenida en la primera quincena de Octubre.**

Así mismo, conviene remarcar que es importante que el estudiante rellene las encuestas de satisfacción de esta asignatura (y de todas) dando sugerencias sobre el desarrollo de la asignatura porque éstas nos ayudan a ir mejorando la docencia en ella.

EQUIPO DOCENTE

Nombre y Apellidos	EMILIO LETON MOLINA (Coordinador/a de asignatura)
Correo Electrónico	emilio.leton@dia.uned.es
Teléfono	91398-9473
Facultad	ESCUELA TÉCN.SUP INGENIERÍA INFORMÁTICA
Departamento	INTELIGENCIA ARTIFICIAL

Nombre y Apellidos	JAVIER OLIVARES ROMERO
Correo Electrónico	jolivares@dia.uned.es
Teléfono	91398-6834
Facultad	ESCUELA TÉCN.SUP INGENIERÍA INFORMÁTICA
Departamento	INTELIGENCIA ARTIFICIAL

HORARIO DE ATENCIÓN AL ESTUDIANTE

Emilio Letón
•Tfno: 91398-9473, martes lectivos, de 14:30 a 18:30 h.
•C/ Juan del Rosal, 16 Despacho 3.04. ETSI Informática - UNED 28040 Madrid
•Email: emilio.leton@dia.uned.es
•Curso Virtual
Javier Olivares
•Tfno: 91398-6834, miércoles lectivos, de 09:30 a 13:00 h.
•C/ Juan del Rosal, 16 Despacho 3.12. ETSI Informática - UNED 28040 Madrid
•Email: jolivares@dia.uned.es
•Curso Virtual

TUTORIZACIÓN EN CENTROS ASOCIADOS

En el enlace que aparece a continuación se muestran los centros asociados y extensiones en las que se imparten tutorías de la asignatura. Estas pueden ser:
•**Tutorías de centro o presenciales**: se puede asistir físicamente en un aula o despacho del centro asociado.
•**Tutorías campus/intercampus**: se puede acceder vía internet.
Consultar horarios de tutorización de la asignatura 71032021

COMPETENCIAS QUE ADQUIERE EL ESTUDIANTE

La nueva normativa se centra en resultados de aprendizaje.

RESULTADOS DE APRENDIZAJE

RA01– Al finalizar el aprendizaje, el o la estudiante será capaz de aplicar los conocimientos necesarios de álgebra, cálculo, análisis, matemática discreta, lógica y estadística para la resolución de los problemas propios de la inteligencia artificial. TIPO: Habilidades o destrezas.

CONTENIDOS

Tema 1: Simulación estadística de modelos de probabilidad

Este tema comprende el estudio de conceptos básicos de variables aleatorias, principales modelos de probabilidad, transformación de variables aleatorias, conceptos básicos de vectores aleatorios y operaciones con variables aleatorias.

El material de estudio de este tema se detalla en el Manual Didáctico de esta asignatura, que se encuentra en el espacio correspondiente, en la plataforma de cursos virtuales de la UNED.
Antes de comenzar dicho material de estudio se recomienda refrescar los conceptos básicos de variables aleatorias y principales modelos de probabilidad vistos en la asignatura "Fundamentos de estadística para la IA". Se puede consultar además el glosario disponible en el curso virtual.
Los contenidos fundamentales de este bloque son los conceptos básicos de variables aleatorios y los contenidos de mayor dificultad son los relativos a operaciones con variables aleatorias.

Tema 2: Índices de riesgo

Este tema explica las medidas básicas de riesgo en una tabla de contingencia (diferencia de riesgos, riesgo relativo y odds ratio). Los conceptos de este tema son básicos para poder interpretar la regresión logística del tema 7.
El material de estudio de este tema se detalla en el Manual Didáctico de esta asignatura, que se encuentra en el espacio correspondiente, en la plataforma de cursos virtuales de la UNED.
Antes de comenzar dicho material de estudio se recomienda refrescar los conceptos básicos de intervalos de confianza y contraste de hipótesis vistos en la asignatura "Fundamentos de estadística para la IA". Se puede consultar además el glosario disponible en el curso virtual.
Los contenidos fundamentales de este bloque son los relativos al concepto de tabla de contingencia y los contenidos de mayor dificultad son los relativos al cálculo de los intervalos de confianza de los índices de riesgo.

Tema 3: Técnicas diagnósticas + Trabajo teórico

En el caso de que una variable explicativa se utilice para diagnosticar aparecen nuevos conceptos estadísticos como: sensibilidad y especificidad, valores predictivos, likelihood ratios y curva ROC. Este tema es fundamental para entender el comportamiento del análisis discriminante y de la regresión logística en los temas 6 y 7 respectivamente.
El material de estudio de este tema se detalla en el Manual Didáctico de esta asignatura, que se encuentra en el espacio correspondiente, en la plataforma de cursos virtuales de la UNED.
Los contenidos fundamentales de este bloque son los de sensibilidad y especificidad y los contenidos de mayor dificultad son los relativos a curva ROC.

Tema 4: Regresión lineal

En este tema se estudia el modelo básico de regresión lineal que modeliza una respuesta cuantitativa. Directamente se estudia el caso de varias variables explicativas, por lo que este tema requiere una especial concentración. Este tema es muy importante en el desarrollo del curso.
El material de estudio de este tema se detalla en el Manual Didáctico de esta asignatura, que se encuentra en el espacio correspondiente, en la plataforma de cursos virtuales de la UNED.
En este tema se utiliza álgebra matricial que se repasa en el curso virtual a través de pequeños ejercicios que se plantean en el foro.
Los contenidos fundamentales de este bloque son los relativos a las estimaciones puntuales de los coeficientes de la regresión lineal y los contenidos de mayor dificultad son los relativos al cálculo matricial involucrado.

Tema 5: Variantes del modelo de regresión lineal + Práctica 1

En este tema se tratan algunas variantes de modelos de regresión lineal como ridge regression, LASSO, PCR o PLS. Para entender este tema es fundamental haber entendido perfectamente el tema anterior.
El material de estudio de este tema se detalla en el Manual Didáctico de esta asignatura, que se encuentra en el espacio correspondiente, en la plataforma de cursos virtuales de la UNED.
Los contenidos fundamentales de este bloque son los relativos a entender las diferencias entre las variantes y los contenidos de mayor dificultad son los relativos a entender las mejoras que proporcionan cada variante.

Tema 6: Análisis discriminante

En este tema se estudia el primero de los modelos básicos de clasificación que modeliza una respuesta cualitativa. Este tema es importante desde un punto de vista histórico para entender los modelos de clasificación.
El material de estudio de este tema se detalla en el Manual Didáctico de esta asignatura, que se encuentra en el espacio correspondiente, en la plataforma de cursos virtuales de la UNED.
Los contenidos fundamentales de este bloque son los relativos a las estimaciones puntuales de los coeficientes del análisis discriminante y los contenidos de mayor dificultad son los relativos al cálculo matricial involucrado.

Tema 7: Regresión logística

En este tema se estudia el segundo de los modelos básicos de clasificación que modeliza una respuesta cualitativa. Este tema presenta uno de los modelos estadísticos más usados en la práctica. Además, se ponen en práctica todos los conceptos trabajados en temas anteriores que han servido de preparación a éste.
El material de estudio de este tema se detalla en el Manual Didáctico de esta asignatura, que se encuentra en el espacio correspondiente, en la plataforma de cursos virtuales de la UNED.
Los contenidos fundamentales de este bloque son los relativos a las estimaciones puntuales de los coeficientes de la regresión logística y los contenidos de mayor dificultad son los relativos al cálculo iterativo en su cálculo involucrado.

Tema 8: Modelos lineales generalizados

En este tema se tratan algunos de los modelos lineales generalizados más comunes en la práctica como el modelo de regresión de Poisson.
El material de estudio de este tema se detalla en el Manual Didáctico de esta asignatura, que se encuentra en el espacio correspondiente, en la plataforma de cursos virtuales de la UNED.
Los contenidos fundamentales de este bloque son los relativos a entender las diferencias entre modelos lineales generalizados y los contenidos de mayor dificultad son los relativos a entender las mejoras que proporcionan cada modelo.

Tema 9: Evaluación y selección de modelos + Práctica 2

En este tema se describen los métodos principales de evaluación de modelos y cómo usar dichos métodos para elegir entre varios modelos.
El material de estudio de este tema se detalla en el Manual Didáctico de esta asignatura, que se encuentra en el espacio correspondiente, en la plataforma de cursos virtuales de la UNED.
Los contenidos fundamentales de este bloque son los relativos a los distintos índices de evalaución y los contenidos de mayor dificultad son los relativos al manejo de la multicolinealidad en los modelos.

METODOLOGÍA

Esta asignatura ha sido diseñada para la enseñanza a distancia. Por tanto, el sistema de enseñanza-aprendizaje estará basado en gran parte en el estudio independiente o autónomo del estudiante. Para ello, **el estudiante contará con un Manual Didáctico que permitirá su trabajo autónomo**. Asimismo, mediante la plataforma virtual de la UNED existirá un contacto continuo entre el equipo docente y los estudiantes, así como una interrelación entre los propios estudiantes a través de los foros, importantísimo en la enseñanza no presencial.
Las **actividades formativas** para el estudio de la asignatura son las siguientes:

•Aprendizaje autónomo de los contenidos teóricos: 50h.

•Aprendizaje autónomo mediante la realización de problemas: 26h.

•Aprendizaje con apoyo docente síncrono: 16h.

•Aprendizaje con apoyo docente asíncrono: 4h.

•Evaluación formativa: 14h.

•Prácticas de laboratorio no presenciales asíncronas: 40h.

En las actividades anteriores se contemplan las siguientes tareas con puntuación que complementan a la evaluación presencial:

1. Participación en el foro

La puntuación de la participación en el foro será de 0 a 10. Esta tarea supone una dedicación de 12 horas de evaluación formativa: es la prueba de evaluación continua (PEC). Por cada participación relevante, el equipo docente dará un punto al alumno. Se entiende por participación relevante cuando se contesta de manera acertada la pregunta de otro alumno o cuando se comunica un aspecto interesante relacionado con la asignatura. No se considera relevante preguntar simplemente una pregunta con una duda en el foro.
Para favorecer que cualquier estudiante pueda participar en el foro, el equipo docente irá también haciendo preguntas cada semana, por lo que todo el mundo podrá optar a puntuar en este apartado, incluso aunque se incorpore tarde al estudio de la asignatura. No se tendrán en cuenta contestaciones repetidas de otros alumnos. Señalar, por último, que la experiencia demuestra que no es difícil conseguir la máxima puntuación y que no hay que agobiarse por conseguirla en las primeras semanas: ¡¡hay tiempo suficiente!!
El último día que se considera para poder puntuar en la participación en el foro es el 13 de Enero. En la plataforma virtual estará publicada una tarea desde finales de Octubre para que el estudiante rellene la fecha de su última aportación relevante. Esto permite que el equipo docente pueda decirle su nota con anterioridad a examinarse y en caso de que hubiera tiempo, incluso podría mejorarla. Para aquellos estudiantes que no puedan seguir el ritmo participativo del foro (a pesar de los beneficios que ello conlleva para el aprendizaje) se les ofrecerá una tarea de forma alternativa para sustituir dicha participación: habrá que elegir una u otra. Esta opción se publicará a finales de Octubre con fecha de entrega también el 13 de Enero.
Conviene remarcar que el equipo docente sólo evaluará la participación en el foro durante el

semestre que se imparte la asignatura, aunque la nota se mantendrá para Septiembre. No obstante, en la convocatoria de Septiembre habrá también una tarea equivalente a la participación en el foro para aquellos que no hubieran participado durante el periodo lectivo (primer cuatrimestre), que se publicará a finales de Abril.

2. Trabajo teórico

La puntuación del trabajo teórico será de 0 a 10. Esta tarea supone una dedicación de 10 horas.

El último día para entregar el trabajo teórico es el 13 de Enero. En la plataforma virtual estará publicada una tarea desde finales de Octubre para que el estudiante pueda entregar dicho trabajo.

El trabajo teórico para la convocatoria de Septiembre se publicará a finales de Abril y tendrá como fecha de entrega los primeros días de Septiembre.

3. Prácticas de laboratorio no presenciales asíncronas

La puntuación de cada práctica será de 0 a 10. Es obligatorio aprobar las prácticas para aprobar la asignatura. Esta tarea supone una dedicación de 40 horas. Hay dos prácticas.

La primera práctica se publica a mediados de Noviembre y su fecha de entrega es en los primeros días de Diciembre. La segunda práctica se publica a mediados de Diciembre y su fecha de entrega es en los primeros días de Enero. No obstante, se dará una semana de prórroga en cada práctica aunque en ese caso sólo se podrá optar a un 9. Adicionalmente, para aquellos estudiantes que no puedan seguir el ritmo de entrega de las prácticas (a pesar de los beneficios que ello conlleva para el aprendizaje) o que no aprueben las prácticas, se establecerán otras prácticas con fecha aproximada de entrega a finales de Enero. Esas nuevas prácticas se puntúan sobre 10 y no tendrán posibilidad de prórroga.

Las prácticas para la convocatoria de Septiembre se publicarán a finales de Abril y tendrán como fecha de entrega los primeros días de Septiembre, puntuándose sobre 10 y sin posibilidad de prórroga.

SISTEMA DE EVALUACIÓN

TIPO DE PRUEBA PRESENCIAL

Tipo de examen	Examen de desarrollo
Preguntas desarrollo	7
Duración del examen	120 (minutos)

Material permitido en el examen

Todo tipo de material escrito y calculadora programable.

Criterios de evaluación

El examen (tanto en primera como en segunda semana) consta de 6 preguntas de verdadero/falso para justificar y una pregunta de desarrollo. En el examen se indica que las preguntas no acertadas se puntúan con un 0, las acertadas sin justificación o con justificación incorrecta puntúan con un 0.1 y que las acertadas con justificación completa mediante una demostración o un contraejemplo se puntúan hasta con un 1.5. La pregunta de desarrollo puntúa un máximo de un punto.

% del examen sobre la nota final	40
Nota del examen para aprobar sin PEC	
Nota máxima que aporta el examen a la calificación final sin PEC	4
Nota mínima en el examen para sumar la PEC	0

Comentarios y observaciones

El examen podrá contener preguntas extraidas del texto base y del material alojado en el curso virtual: mini-libros, pácticas y las preguntas más frecuentes (P+F) que el equipo docente extrae de las interacciones en los foros.

PRUEBAS DE EVALUACIÓN CONTINUA (PEC)

¿Hay PEC? Si

Descripción

La evaluación continua en esta asignatura se hace a través de la participación en el foro. La información detallada se encuentra en el apartado de "Metodología"

Criterios de evaluación

Los criterios detallados se encuentran en el apartado de "Metodología".

Ponderación de la PEC en la nota final	La ponderación de la participación en el foro es de un 10%.
Fecha aproximada de entrega	Foro: 13-Ene

Comentarios y observaciones

No es obligatoria presentar la tarea del foro, aunque si no se hace se obtiene un 0 en dicha tarea. Tanto en Febrero como en Septiembre hay una tarea equivalente al foro. En Febrero habrá que elegir una u otra, en Septimebre sólo está la tarea equivalente al foro y se describirá en el curso virtual a finales de Abril.

OTRAS ACTIVIDADES EVALUABLES

¿Hay otra/s actividad/es evaluable/s? Si

Descripción

Otras tareas evaluables son el trabajo teórico y las prácticas de laboratorio no presenciales asíncronas. La información detallada se encuentra en el apartado de "Metodología"

Criterios de evaluación

Los criterios detallados se encuentran en el apartado de "Metodología".

Ponderación en la nota final	La ponderación del trabajo teórico es de un 10% y la de cada práctica de un 20%.

Fecha aproximada de entrega	Trabajo teórico: 13-Ene; Práctica 1: 13-Dic; Práctica 2: 13-Ene

Comentarios y observaciones

No es obligatorio presentar la tarea del trabajo teórico, aunque si no se hace se obtiene un 0 en dicha tarea. En Septiembre habrá otro trabajo teórico que se describirá en el curso virtual a finales de Abril.

Es obligatorio presentar y aprobar (nota mayor igual a 5) las prácticas de laboratorio no presenciales asíncronas. En la convocatoria de Septiembre habrá otras prácticas que se describirán en el curso virtual a finales de Abril.

¿CÓMO SE OBTIENE LA NOTA FINAL?

Nota examen (Ex): 0-10.
Nota práctica 1 (P1): 0-10.
Nota práctica 2 (P2): 0-10.
Nota participación curso virtual (F): 0-10.
Nota trabajo teórico (T): 0-10.
Nota final NF=0,4*Ex+0,2*P1+0,2*P2+0,1*F+0,1*T

BIBLIOGRAFÍA BÁSICA

En el espacio de esta asignatura, en la plataforma de cursos virtuales de la UNED, se encuentra un Manual Didáctico específico donde se detalla todo el material de estudio. En algunos temas también se utilizará el siguiente libro:

- **James, G., Witten, D., Hastie, T., &Tibshirani, R. (2021). An introduction to statistical learning with Applications in R. New York: Springer. El libro está disponible en abierto desde el enlace https://www.statlearning.com/.**

Existe una versión en Python, pero se utiliza la versión en R. En concreto, las páginas del libro ISLR que se van a trabajar son: 71-104, 110-121, 133-153, 164-180, 185-189, 197-219 y 224-282.

El hecho de utilizar R ayuda a que los estudiantes amplíen conocimientos de un nuevo lenguaje de programación muy utilizado en Estadística, complementando así la formación que se recibe en otras asignaturas del Grado. Además, se da continuidad a la asignatura de "Fundamentos de estadística para la IA" donde también se utiliza R.

BIBLIOGRAFÍA COMPLEMENTARIA

La bibliografía complementaria de esta asignatura es:

- Hosmer Jr, D. W., Lemeshow, S., &Sturdivant, R. X. (2013). Applied logistic regression. John Wiley &Sons.
- Letón, E., &Pedromingo, A. (2001). Introducción al análisis de datos en meta-análisis. Díaz de Santos, Madrid.

•Peña, D. (2013). Análisis de datos multivariantes. McGraw-Hill España.
•Peña, D. (2014). Fundamentos de estadística. Alianza Editorial, Madrid.
•Pepe, M. S. (2003). The statistical evaluation of medical tests for classification and prediction. Oxford University Press.

Los cuatro primeros de la bibliografía complementaria se encuentran disponibles en la red de bibliotecas de la UNED, se puede consultar en el enlace página web de la Biblioteca UNED.

RECURSOS DE APOYO Y WEBGRAFÍA

En esta asignatura se utiliza el programa estadístico R, disponible en la página web de R.

IGUALDAD DE GÉNERO

En coherencia con el valor asumido de la igualdad de género, todas las denominaciones que en esta Guía hacen referencia a órganos de gobierno unipersonales, de representación, o miembros de la comunidad universitaria y se efectúan en género masculino, cuando no se hayan sustituido por términos genéricos, se entenderán hechas indistintamente en género femenino o masculino, según el sexo del titular que los desempeñe.