



## FICHA IDENTIFICATIVA

## Datos de la Asignatura

|                 |                                    |
|-----------------|------------------------------------|
| Código          | 36518                              |
| Nombre          | Predicción con Datos Transversales |
| Ciclo           | Grado                              |
| Créditos ECTS   | 6.0                                |
| Curso académico | 2023 - 2024                        |

## Titulación(es)

| Titulación                                               | Centro               | Curso | Periodo             |
|----------------------------------------------------------|----------------------|-------|---------------------|
| 1332 - Grado en Inteligencia y Analítica de Negocios/BIA | Facultad de Economía | 2     | Primer cuatrimestre |

## Materias

| Titulación                                               | Materia                                           | Caracter    |
|----------------------------------------------------------|---------------------------------------------------|-------------|
| 1332 - Grado en Inteligencia y Analítica de Negocios/BIA | 24 - Herramientas y Técnicas de Análisis de Datos | Obligatoria |

## Coordinación

| Nombre                           | Departamento            |
|----------------------------------|-------------------------|
| GOERLICH GISBERT, FRANCISCO JOSE | 10 - Análisis Económico |

## RESUMEN

“Predicción con Datos Transversales” es una asignatura de formación básica adscrita a las áreas de Métodos Cuantitativos para la Economía y la Empresa y Fundamentos del Análisis Económico que se imparte en el primer semestre del segundo curso del Grado de INTELIGENCIA Y ANALÍTICA DE NEGOCIOS con una carga lectiva total de 6 créditos ECTS.

En el marco de una titulación claramente orientada a formar profesionales de empresa con profundos conocimientos del análisis y proceso de grandes volúmenes de información esta es la primera asignatura, de un conjunto de 4, dedicada a técnicas predictivas, y lo hace en el marco de datos de corte transversal y a partir de los modelos estadísticos más sencillos posibles. Otras estructuras de datos y modelos más complejos serán cubiertas en asignaturas posteriores.

La asignatura tiene una orientación eminentemente práctica y computacional. Por esta razón los desarrollos teóricos serán presentados de forma breve y esquemática, haciendo especial hincapié en la aplicabilidad de los mismos y la intuición de los procedimientos –las clases teóricas representan el 25% del total de horas lectivas–. Las prácticas se desarrollarán sobre ejemplos con datos reales y *software* específico y el énfasis se pondrá en la replicabilidad por parte del estudiante de los resultados, así como



en la toma de decisiones en el proceso de construcción de modelos predictivos sencillos.

## CONOCIMIENTOS PREVIOS

### Relación con otras asignaturas de la misma titulación

No se han especificado restricciones de matrícula con otras asignaturas del plan de estudios.

### Otros tipos de requisitos

Se asume que para cursar esta asignatura con éxito el estudiante tiene un nivel de matemáticas básico (los conocimientos que corresponden a primero y segundo de bachillerato en la rama de ciencias o ciencias sociales) y está familiarizado con los contenidos de la materia Análisis Exploratorio de Datos y Azar, Incertidumbre e Inferencia cursadas con anterioridad. Al mismo tiempo se supone que ha adquirido algunas de las competencias básicas de las programadas con anterioridad sobre gestión de la información, programación y uso de las TICs.

## COMPETENCIAS

### 1332 - Grado en Inteligencia y Analítica de Negocios/BIA

- Que los estudiantes hayan demostrado poseer y comprender conocimientos en un área de estudio que parte de la base de la educación secundaria general, y se suele encontrar a un nivel que, si bien se apoya en libros de texto avanzados, incluye también algunos aspectos que implican conocimientos procedentes de la vanguardia de su campo de estudio.
- Que los estudiantes sepan aplicar sus conocimientos a su trabajo o vocación de una forma profesional y posean las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y la resolución de problemas dentro de su área de estudio.
- Que los estudiantes tengan la capacidad de reunir e interpretar datos relevantes (normalmente dentro de su área de estudio) para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética.
- Que los estudiantes puedan transmitir información, ideas, problemas y soluciones a un público tanto especializado como no especializado.
- Que los estudiantes hayan desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía.
- Conocimiento de materias básicas que capacite para el aprendizaje de nuevos métodos y tecnologías, y que le dote de versatilidad para adaptarse a nuevas situaciones en los ámbitos académico y profesional.
- Capacidad para resolver problemas, y para comunicar y transmitir conocimientos, habilidades y destrezas, comprendiendo la responsabilidad ética, igualitaria y profesional de la actividad de la Inteligencia y Analítica de Negocios.



- Capacidad para tomar decisiones de forma autónoma en entornos digitales caracterizados por la abundancia y dinamismo de los datos.
- Conocer y saber utilizar adecuadamente los diferentes métodos cuantitativos y cualitativos apropiados para razonar analíticamente, evaluar resultados y predecir magnitudes económicas y financieras.
- Capacidad para aplicar métodos analíticos y matemáticos para el análisis de los problemas económicos y empresariales.
- Capacidad para planificar, organizar, controlar y evaluar la puesta en marcha de estrategias empresariales.
- Capacidad de análisis y síntesis
- Capacidad para analizar y buscar información proveniente de fuentes diversas.
- Capacidad de aprendizaje autónomo.
- Capacidad para utilizar las TIC, tanto en el ámbito de estudio como en el desarrollo profesional.
- Capacidad para definir, resolver y exponer de forma sistémica problemas complejos.
- Capacidad para trabajar en equipo, con el compromiso por la calidad, la ética, la igualdad entre personas y la responsabilidad social.
- Aplicar métodos y técnicas de análisis, síntesis y representación gráfica mediante programas informáticos.
- Expresar las situaciones de incertidumbre y azar utilizando lenguajes matemáticos, sintéticos y gráficos.
- Reorganizar y reestructurar variables y bases de datos.
- Aplicar minería de datos mediante programas informáticos.
- Utilizar software para resolver problemas con incertidumbre.
- Distinguir entre los enfoques explicativo y predictivo en el análisis de datos y en los negocios.
- Predecir utilizando software adecuado al manejo de series temporales.
- Aplicar muestreos probabilísticos y no probabilísticos.
- Utilizar software para recolectar y analizar datos de encuestas.

## **RESULTADOS DE APRENDIZAJE**

El objetivo del curso es que el estudiante sea capaz de resolver problemas predictivos mediante modelización estadística en un contexto de datos de corte transversal.

Se pretende que el estudiante domine las técnicas más sencillas, partiendo de los modelos de regresión simple y algunas de sus extensiones básicas: variables cualitativas, modelos no lineales, modelos lineales generalizados, modelos locales.



Igualmente se persigue el objetivo de que el estudiante maneje estos métodos con herramientas informáticas potentes integrables en los análisis de situaciones con grandes volúmenes de datos. De esta forma los aspectos computacionales son un aspecto totalmente esencial del curso.

## DESCRIPCIÓN DE CONTENIDOS

### 1. Introducción

Estructuras de datos y tipos de variables.  
Variables respuesta y predictores.  
Modelización o aprendizaje estadístico.

### 2. Modelo de regresión lineal simple

Estimación.  
Mínimos Cuadrados Ordinarios.  
Bondad del ajuste.

### 3. Método de los Vecinos más Cercanos

Planteamiento general

### 4. Modelo de regresión lineal múltiple

Selección de regresores.  
Criterios de información.  
Predicción.

### 5. Medición de la precisión (accuracy) en un contexto predictivo.

Conjunto de entrenamiento (training set) y conjunto de contraste (test set).  
El trade-off sesgo-varianza.

### 6. Modelos no-lineales

Transformación de regresores.  
Mínimos Cuadrados No-lineales.

**7. Predictores cualitativos - Factores**

Tratamiento de la información cualitativa como predictores en el modelo de regresión.

**8. Modelos lineales generalizados**

Output binario.

Modelo lineal de probabilidad.

Modelo Logit/Probit.

Errores de clasificación en un contexto binario.

**9. Métodos de re-muestreo**

Validación cruzada (cross-validation)

Bootstrap

**10. Regularización**

Step-wise regression

Ridge-regression

Lasso-regression

**11. Modelos de regresión locales**

Splines

Regresión local (lowess)

GAM

**VOLUMEN DE TRABAJO**

| ACTIVIDAD                                      | Horas         | % Presencial |
|------------------------------------------------|---------------|--------------|
| Prácticas en aula informática                  | 45,00         | 100          |
| Clases de teoría                               | 15,00         | 100          |
| Elaboración de trabajos individuales           | 10,00         | 0            |
| Estudio y trabajo autónomo                     | 20,00         | 0            |
| Preparación de actividades de evaluación       | 20,00         | 0            |
| Preparación de clases de teoría                | 10,00         | 0            |
| Preparación de clases prácticas y de problemas | 10,00         | 0            |
| Resolución de casos prácticos                  | 20,00         | 0            |
| <b>TOTAL</b>                                   | <b>150,00</b> |              |





## METODOLOGÍA DOCENTE

El desarrollo de la asignatura se estructura fundamentalmente en torno a las sesiones teóricas y prácticas diferenciadas, siendo las teóricas de 1 hora a la semana (25%) y las prácticas de 3 horas a la semana (75%). La metodología, por tanto, enfatiza los aspectos más prácticos y computacionales de la asignatura.

En las sesiones teóricas se presentarán los fundamentos de la modelización o aprendizaje estadístico con especial énfasis en los aspectos más intuitivos, y relegando los aspectos más teóricos y formales a lecturas complementarias.

En las sesiones prácticas se pondrán en práctica los aspectos teóricos desde un punto de vista computacional, utilizando *software* específico para ello, en concreto *R*, que se ha constituido en los últimos años en el *software* de referencia en cuestiones estadísticas y de análisis de datos. El alumno deberá manejar estas herramientas para resolver cuestiones prácticas bajo la tutela del profesor. De esta forma se otorga cierto papel a los aspectos más prácticos del manejo de las herramientas de informática, y que es absolutamente esencial en la sociedad moderna.

El método docente predominante en las clases teóricas será la clase magistral participativa. Esta metodología permite dirigir de forma organizada los grupos grandes de alumnos ofreciendo las ventajas de una clase magistral sin limitar por ello, la participación de los alumnos y la interacción profesor-estudiante. Se intentará fomentar la participación y la discusión en la clase, con el fin de ofrecer a los estudiantes una implicación directa con el contenido.

En las sesiones prácticas, de tres horas de duración, el profesor propondrá a los alumnos ejercicios, fundamentalmente a partir de datos reales, para la resolución de problemas o estudios de casos que deberán resolver con aplicación de las técnicas mencionadas en las clases teóricas así como la utilización de programas informáticos adecuados, realizando si es pertinente, presentaciones orales o debates..., individualmente y/o en equipo. En las clases prácticas se propondrán proyectos y situaciones que los estudiantes deberán resolver entregando en tiempo y forma los resultados que se determinen.

## EVALUACIÓN

La evaluación constará de dos partes:

- **Evaluación continua** de cada alumno, basada en la resolución de ejercicios frecuentes, tanto fuera de aula como en el aula y la entrega de ejercicios. La evaluación continua supondrá el **80%** de la nota final.
- Una **prueba final** que consistirá en la resolución de ejercicios con código en la fecha prevista por la faultad para la realización de exámenes. La prueba final supondrá el **20%** de la nota final.

La evaluación continua será no recuperable.



## REFERENCIAS

### Básicas

- James, G.; Witten, D.; Hastie, T. & Tibshirani, R. (2013) An Introduction to Statistical Learning. Springer Texts in Statistics. Springer. New York.
- Kuhn, M. & Johnson, K. (2013) Applied Predictive Modeling. Springer. New York.
- Hanck, C.; Arnold, M.; Gerber, A. & Schmelzer, M. (2019) Introduction to Econometrics with R. UNIVERSITÄT DUISBURG ESSEN. Open-Minded.
- Berk, R. A. (2016) Statistical Learning from a Regression Perspective. Springer. New York.

### Complementarias

- Hastie, T.; Tibshirani, R. & Friedman, J. (2008) The Elements of Statistical Learning. Data Mining, Inference and Prediction. Springer Texts in Statistics. 2nd edition. Springer. New York.