

# Minería de datos

**Créditos: 3 ECTS**

## **Estructura**

**Módulo :** Fundamentos

**Materia:** Minería de datos

## **Contenido:**

1. Introducción a la minería de datos: Planteamiento de una problemática de negocio
2. Análisis descriptivo
3. Técnicas de muestreo
4. Técnicas de manipulación de datos: transformación de variables, tratamiento de outliers, imputación de missings
5. Técnicas de modelización supervisada: Árboles de clasificación/regresión, regresión logística, Redes neuronales, Ensamblaje de modelos: métodos de bagging y boosting
6. Métricas de validación de resultados: matriz de confusión, gráficos de respuesta y lift, curva roc, etc.

## **Resultados del aprendizaje:**

El alumno adquirirá los conocimientos teóricos necesarios para abordar diferentes problemáticas de negocio a través del tratamiento de datos y el ajuste de modelos de carácter predictivo.

También conseguirá manejar con destreza la herramienta Enterprise Miner de SAS

## **Competencias:**

Competencias básicas

CB6 - Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación

CB7 - Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio

CB9 - Que los estudiantes sepan comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades

CB10 - Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.

#### Competencia generales

CG1 - Aprender a aplicar los conocimientos adquiridos y a explotar su potencial para la resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) en el tratamiento estadístico-computacional de la información.

CG2 - Elaborar adecuadamente y con originalidad argumentos motivados y proyectos de trabajo, redactar planes, así como formular hipótesis y conjeturas razonables en su área de especialización.

CG3 - Integrar los conocimientos adecuados y enfrentarse a la complejidad de emitir juicios en función de criterios, de normas externas o de reflexiones personales justificadas

CG4 - Comunicar y presentar públicamente ideas, procedimientos o informes de investigación, así como asesorar a personas u organizaciones en el tratamiento estadístico-computacional de la información. La presentación de estas ideas debe transmitir de forma clara y precisa las conclusiones de forma que sean entendidas tanto por el especialista como por el profano en temas estadístico-computacionales.

CG7 - Saber abstraer en un modelo matemático las propiedades y características esenciales de un problema real reconociendo su rango de aplicabilidad y limitaciones.

#### Competencia transversales

CT1 - Saber aplicar sus conocimientos a su trabajo o vocación de una forma profesional y poseer las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y en la resolución de problemas y estudio de casos. Esto implica, más concretamente: Integrar creativamente conocimientos y aplicarlos a la resolución de problemas complejos, perseguir objetivos de calidad en el desarrollo de su actividad profesional, adquirir capacidad para la toma de decisiones y de dirección de recursos humanos, ser capaz de mostrar creatividad, iniciativa y espíritu emprendedor para afrontar los retos de su actividad, valorar la importancia de los métodos estadístico-computacionales en el contexto industrial, económico, administrativo, medio ambiental y social.

CT2 - Tener la capacidad de reunir e interpretar datos relevantes para emitir juicios que incluyan una reflexión sobre temas relevantes de índole científica, tecnológica y empresarial. Demostrar razonamiento crítico y gestionar información científica y técnica de calidad, bibliografía, bases de datos especializadas y recursos accesibles a través de Internet.

#### Competencias específicas

CE1 - Adquisición de una formación sólida y rigurosa en temas avanzados de Estadística, Matemática computacional, Modelos estocásticos y Metodología de la toma de Decisiones aplicadas al tratamiento de la Información.

CE2 - Capacidad para planificar la resolución de un problema en función de las herramientas de que se disponga y, en su caso, de las restricciones de tiempo y recursos.

CE3 - Capacidad para utilizar aplicaciones informáticas estadísticas, de cálculo numérico y simbólico, visualización gráfica, optimización u otras para resolver problemas con un elevado grado de complejidad.

CE4 - Desarrollar habilidades de aprendizaje en Estadística Computacional y Matemáticas, así como en sus respectivas aplicaciones, que permitan al alumno continuar estudiando y profundizando en la materia de modo autónomo, así como el desarrollo profesional con un alto grado de independencia.

CE5 - Resolver problemas y casos reales planteados en el tratamiento estadístico-computacional de la información generada en los ámbitos de la ciencia, la tecnología y la sociedad mediante habilidades de modelización matemática, estimación y computación.

CE7 - Capacidad de utilización de herramientas de búsqueda de recursos bibliográficos así como manejo, gestión y análisis de grandes bases de datos.

#### **Bibliografía:**

- Bishop, C.M. (2006), "Pattern Recognition and Machine Learning", Springer
- Hastie, T. Tibshirani, R. Friedman, J. (2001), "The Elements of Statistical Learning", Springer
- Hair-Anderson-Tatham-Black (1999), "Análisis multivariante", Prentice Hall
- Pérez, C. (2001), "El sistema estadístico SAS", Prentice Hall

#### **Metodología docente**

- Método expositivo
- Estudio de casos
- Prácticas de ordenador

#### **Evaluación:**

Presentación de un trabajo en el que contraste, sobre un problema real, la bondad de las diferentes técnicas de modelización aprendidas

#### **Profesorado:**

Nombre: Daniel Vélez Serrano

Despacho : 410

Facultad de Matemáticas

Teléfono: 913944729

Correo electrónico: [danielvelezserrano@mat.ucm.es](mailto:danielvelezserrano@mat.ucm.es)