

Ciencia de Datos

Código: 104540
Créditos ECTS: 6

Titulación	Tipo	Curso	Semestre
2503743 Gestión de Ciudades Inteligentes y Sostenibles	OB	2	2

Contacto

Nombre: Dimosthenis Karatzas

Correo electrónico: Dimosthenis.Karatzas@uab.cat

Uso de idiomas

Lengua vehicular mayoritaria: español (spa)

Algún grupo íntegramente en inglés: No

Algún grupo íntegramente en catalán: No

Algún grupo íntegramente en español: No

Prerequisitos

Haber cursado las asignaturas Informática, Matemáticas, Programación de aplicaciones en Internet, de primer curso, y Bases de Datos de segundo curso.

Objetivos y contextualización

Esta asignatura debe permitir al estudiante conocer las tecnologías existentes y las diversas formas de abordar la gestión y análisis de los datos generados por la ciudad a diario.

Los estudiantes aprenderán técnicas de visualización, análisis y modelaje de datos que les permitirán generar nuevo conocimiento y percepciones a partir de los datos de la ciudad.

Competencias

- 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 puedan transmitir información, ideas, problemas y soluciones a un público tanto especializado como no especializado.
- 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.
- Resolver problemas de gestión urbana utilizando conocimientos, metodologías y procedimientos de diseño e implementación de aplicaciones informáticas para diferentes tipos de entornos (web, móvil, nube) y con diferentes paradigmas.
- Trabajar cooperativamente, en entornos complejos o inciertos y con recursos limitados, en un contexto multidisciplinar, asumiendo y respetando el rol de los diferentes miembros del equipo.

Resultados de aprendizaje

1. Aplicar técnicas de toma de decisiones automatizada.
2. 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.
3. Que los estudiantes puedan transmitir información, ideas, problemas y soluciones a un público tanto especializado como no especializado.
4. 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.
5. 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.
6. Trabajar cooperativamente, en entornos complejos o inciertos y con recursos limitados, en un contexto multidisciplinar, asumiendo y respetando el rol de los diferentes miembros del equipo.

Contenido

- Preparación de datos
 - Visualización de datos
 - Normalización
 - Valores desconocidos
 - Reducción de dimensionalidad
 - Selección de características
- Clasificación (técnicas supervisadas)
 - Probabilidades, Clasificador Naive Bayes
 - Árboles de decisión y "random forests"
 - Clasificación jerárquica
- Generación de conocimiento (técnicas no supervisadas)
 - Reglas de asociación
 - Sistemas de recomendación

Metodología

La ciencia de datos se define por los tipos de problemas que intenta solucionar, por lo tanto, será la tipología de problemas la que dirigirá la organización de todos los contenidos.

Se harán tres tipos de sesiones:

Clases de teoría: Serán clases de tipo magistrales en las que el objetivo es que el profesor explique los contenidos teóricos de la asignatura. Para cada uno de los temas se plantea un tipo de problema que se pretende solucionar y a partir de ahí se van planteando los sub-objetivos y se explican las correspondientes soluciones algorítmicas a cada uno de ellos.

Clases de problemas: Serán clases que faciliten la interacción. En estas clases se persigue reforzar la comprensión de los temas de la teoría planteando casos prácticos que requieran el diseño de una solución en la que se utilicen los métodos vistos en las clases de teoría.

Laboratorio de prácticas: Serán sesiones en las que se llevarán a cabo diferentes tipos de actividades relacionadas con la realización en equipo de dos proyectos de manera secuencial. En las sesiones de prácticas se hará la presentación de los proyectos a resolver y una serie de actividades que se llevarán a cabo en equipos de estudiantes en modo de trabajo colaborativo. Se deberá abordar la identificación del problema, la discusión del diseño, el reparto y la organización del trabajo a realizar, el desarrollo de la solución y la presentación de los resultados al profesor y al resto de los estudiantes.

Toda la información de la asignatura y los documentos relacionados que los estudiantes necesiten se encontrarán en el campus virtual.

El profesor dará comentarios individualizados para cada una de las actividades entregadas por los estudiantes. Se establecerá un sistema de tutorías y de consultas fuera del horario de clase y se fomentará que los estudiantes hagan uso de las mismas.

La competencia transversal T01 se pondrá en práctica mediante el trabajo en equipo y el intercambio colaborativo que supone el desarrollo de los dos proyectos, que se acompaña por actividades supervisadas en el laboratorio de prácticas. La evaluación de los proyectos incluye una presentación oral de cada equipo al resto de la clase, durante cual los alumnos tienen que exponer su trabajo y también explicar la organización del equipo durante el desarrollo del proyecto.

Actividades

Título	Horas	ECTS	Resultados de aprendizaje
Tipo: Dirigidas			
Clase de Problemas	12	0,48	1, 2, 3
Clase de Teoría	26	1,04	2, 3, 4, 5
Tipo: Supervisadas			
Laboratorio de Prácticas	12	0,48	1, 3, 4, 5, 6
Tutorías	5	0,2	1, 3, 4, 5
Tipo: Autónomas			
Dedicación a las prácticas	37	1,48	1, 2, 4, 5, 6
Dedicación a los problemas	12	0,48	1, 2, 4, 5
Lectura y estudio de material	40	1,6	1, 2

Evaluación

Para evaluar el nivel de aprendizaje del estudiante se establece una fórmula que combina el aprendizaje de conocimientos, la capacidad de resolución de problemas y las capacidades de trabajar en equipo, así como de la presentación de los resultados obtenidos.

Nota Final

La nota final se calcula ponderadamente de la siguiente manera y de acuerdo con las diferentes actividades que se llevan a cabo:

$$\text{Nota final} = 0.3 * \text{Nota Teoría} + 0.1 * \text{Nota Problemas} + 0.6 * \text{Nota Prácticas}$$

Se aplicará esta fórmula siempre que la nota de teoría y la nota de prácticas sean mayor que 5. No hay ninguna restricción sobre la nota de problemas. Si haciendo el cálculo de la fórmula sale ≥ 5 , pero no llega al mínimo exigido en alguna de las actividades de evaluación, entonces la nota que se pondrá en el expediente será de 4.5.

Nota de Teoría

La nota de teoría pretende valorar las capacidades individuales del estudiante en cuanto a los contenidos teóricos de la asignatura, esto se hace de forma continuada durante el curso con dos exámenes parciales:

$\text{Nota Teoría} = 0.5 * \text{Nota Examen 1} + 0.5 * \text{Nota Examen 2}$

Examen 1 se hace en la mitad del semestre y sirve para eliminar parte de la materia si se aprueba.

Examen 2 se hace al final del semestre y sirve para eliminar parte de la materia si se aprueba.

Estos exámenes pretenden hacer una evaluación individualizada de las capacidades de cada estudiante para resolver problemas utilizando las técnicas explicadas en clase, así como evaluar el nivel de conceptualización que el estudiante ha hecho de las técnicas vistas. Para poder tener una nota de teoría será necesario que las notas de los exámenes 1 y 2 sean superiores a 4.

Examen de recuperación. En caso de que la nota de teoría no llegue al nivel adecuado para aprobar la teoría, los estudiantes se pueden presentar a un examen de recuperación, destinado a recuperar la parte no superada de la evaluación continua.

Nota de Problemas

La parte de problemas tiene como objetivo conseguir que el estudiante se entrene con los contenidos de la asignatura de manera continuada y que se familiarice directamente en la aplicación de la teoría. Como evidencia de este trabajo se pide la presentación de un portafolio en el que se irán guardando los problemas realizados:

$\text{Nota Problemas} = \text{Evaluación del portafolio}$

Nota de Prácticas

La parte de prácticas tiene un peso esencial en la nota global de la asignatura y pretende que el estudiante se enfrente al problema de diseñar una solución a un problema que se plantea de forma contextualizada y que, por tanto, requiere el diseño de una solución integral, desde la exploración de las técnicas hasta el modelaje de datos. Además, el estudiante debe demostrar sus habilidades para trabajar en equipo y presentar convincentemente los resultados.

Cada uno de los dos proyectos se evalúa a través de su entregable y una presentación oral que harán los estudiantes en clase. La nota se calcula de la siguiente manera:

$\text{Nota Proyecto} = 2/3 * \text{Nota Entregables} + 1/3 * \text{Nota Presentación}$

$\text{Nota Prácticas} = 0.5 * \text{Nota Proyecto 1} + 0.5 * \text{Nota Proyecto 2}$

En caso de no superar alguno de los proyectos de prácticas se permitirá la recuperación de la parte de los entregables (2/3 de la Nota de Proyecto) de los proyectos no superados, pero no la presentación oral.

Notas importantes

Sin perjuicio de otras medidas disciplinarias que se estimen oportunas, y de acuerdo con la normativa académica vigente, la comisión de actividades irregulares por parte del estudiante (por ejemplo, plagiar, copiar, dejar copiar, ...) implicará suspender la actividad correspondiente con un cero (0). Las actividades de evaluación calificadas de esta forma y por este procedimiento no serán recuperables.

En caso de que no se haga ninguna entrega de problemas, no se asista a ninguna sesión de presentación de los proyectos en prácticas de laboratorio y no se haga ningún examen, la nota correspondiente será un "no evaluable". En otro caso, los "no presentados" computan como un 0 para el cálculo de la media ponderada.

Las matrículas de honor se concederán a quienes obtengan una nota superior o igual a 9.0 en cada parte, hasta el 5% de los matriculados según orden descendente de nota final. A criterio del profesorado, también se podrán conceder en otros casos.

Actividades de evaluación

Título	Peso	Horas	ECTS	Resultados de aprendizaje
Entregables de problemas	10	0	0	1, 2, 5
Entregables de proyectos	40	0	0	1, 2, 4, 5, 6
Exámenes	30	5	0,2	2, 3, 4
Presentación de proyectos	20	1	0,04	2, 3, 4, 6

Bibliografía

- Data Science from Scratch: First Principles with Python, Joel Grus, O'Reilly Media, 2015, 1st Ed.
- Pattern Recognition and Machine Learning, Christopher Bishop, Springer, 2011
- Model-Based Machine Learning, J. Winn, C. Bishop, early access: <http://mbmlbook.com/>
- Computational and Inferential Thinking: The Foundations of Data Science, Ani Adhikari and John DeNero, online: <https://ds8.gitbooks.io/textbook/content/>