

DATOS GENERALES DE LA ASIGNATURA

ASIGNATURA: Técnicas Avanzadas de Análisis de Datos

PLAN DE ESTUDIOS:

Máster Universitario en Gestión y Análisis de Grandes Volúmenes de Datos: Big Data (SMA-ABIGDATA)

GRUPO: 2425-01

CENTRO: Escuela Politécnica Superior

CARÁCTER DE LA ASIGNATURA: Obligatorio

ECTS: 6,0

CURSO: 1º

SEMESTRE: 1º Semestre

IDIOMA EN QUE SE IMPARTE:

Castellano, excepto las asignaturas de idiomas que se impartirán en el idioma correspondiente

DATOS DEL PROFESOR

NOMBRE Y APELLIDOS: JORGE CRESPO ÁLVAREZ

EMAIL: jcrespo@uemc.es

TELÉFONO: 983 00 10 00

CV DOCENTE:

Jorge Crespo es Ingeniero Civil por el Instituto Superior Politécnico José Antonio Echeverría de Ciudad de la Habana, Cuba en 2005. En 2011 obtuvo su Doctorado por la Universidad de Cantabria en el Departamento de Transportes y Tecnologías de Proyectos y Procesos, en el cual desarrolló un modelo físico-matemático para la predicción del comportamiento ante incendios de hormigones de alta resistencia por el cual recibió la mención de sobresaliente cum laude. Experto en modelado y simulación computacional, sus tareas docentes han estado mayormente vinculadas a la rama de las matemáticas aplicadas y matemáticas superiores, impartiendo asignaturas como Calculo Diferencial e Integral, Matemática Discreta, Matemática Numérica, Estadística y Bioestadística tanto en entornos presenciales como a distancia. Así mismo Jorge ha dirigido trabajos de Fin de Carrera, Máster y Doctorado en el área de Business Intelligence y Aprendizaje Máquina.

CV PROFESIONAL:

A partir de 2005 Jorge se incorpora al Claustro Docente del Instituto Superior Politécnico José Antonio Echeverría, en La Habana, Cuba, dentro del Departamento de Matemáticas de la Facultad de Ingeniería Civil. En 2006 pasa a formar parte del Grupo de Energía Eólica de Cuba, con la responsabilidad de Ingeniero Supervisor. En 2007 se incorpora al grupo GIDAI de la Universidad de Cantabria para desarrollar su doctorado donde se especializa en modelado y simulación computacional. Ha trabajado como Asesor de Contenido Web para Microsoft. Desde 2014, Jorge desarrolla su carrera profesional como Profesor de las ramas de las Matemáticas de la Universidad Europea del Atlántico, tarea que compagina con la Dirección de Programas de Postgrado, la Coordinación del Máster en Dirección Estratégica en Tecnologías de la Información y la Coordinación del Grado de Ingeniería de Organización Industrial. Es además consultor independiente y colabora o ha colaborado con diversas empresas como Quest Global, Lionbridge, ROMA S.L. y la Universidad Europea Miguel de Cervantes.

CV INVESTIGACIÓN:

En 2007 comenzó su actividad investigadora conjunto con el inicio de su Doctorado. En la actualidad cuenta con más de treinta publicaciones entre revistas, libros y congresos, de las cuales 16 se encuentran recogidas en el Journal of Citation Report. Cuenta con las acreditaciones de Profesor Contratado Doctor y Profesor de Universidad Privada por la Junta de Castilla y León y cuenta con un sexenio de investigación. Desde 2023 es Full Member de Sigma Xi, The Scientific Research Honor Society, la sociedad científica honorífica mas importante del mundo en el campo de las STEM.

DATOS ESPECÍFICOS DE LA ASIGNATURA

DESCRIPTOR DE LA ASIGNATURA:

Supongamos que quiere predecir el precio justo de una vivienda en la que está interesado. Para ello dispone de datos como la localización, superficie, números de plantas o el número de habitaciones de otras viviendas en el vecindario. Su idea es utilizar algún tipo de algoritmo que, basado en dichos datos, le permita realizar una predicción del precio lo más acertada posible.

En esta asignatura se tratarán varios métodos para construir algoritmos de aprendizaje automático que le permitan resolver esta y otras tareas. Además se estudiarán las técnicas requeridas para evaluar el desempeño de los algoritmos desarrollados y como acondicionar los datos para ser explorados y utilizados por parte de los mencionados algoritmos.

En el desarrollo de la asignatura se utilizará Python y scikit-learn, una biblioteca de código abierto para la realización de tareas de aprendizaje automático o Machine Learning.

Prerrequisitos

Conocimientos básicos de programación en Python (nivel grado).

Conocimientos básicos de álgebra y cálculo (nivel grado).

Conocimientos básicos de estadística (nivel grado).

CONTENIDOS DE LA ASIGNATURA:

1. bloque 1

1. Fundamentos de Machine Learning
2. Regularización
3. Selección de características
4. Extracción de características
5. Árboles de clasificación y regresión
6. Métodos de ensembles
7. Máquinas de vectores soporte
8. Métodos kernel

RECURSOS DE APRENDIZAJE:

Los recursos de aprendizaje que se utilizarán en todas las asignaturas de la titulación (salvo las prácticas externas) para facilitar el proceso de enseñanza-aprendizaje, son:

- Campus online de la UEMC (Open Campus)
- Plataforma de Webconference (Zoom work place)

Las comunicaciones con el profesor serán a través de Open Campus vía Mi correo, Tablón o/y Foro.

COMPETENCIAS Y RESULTADOS DE APRENDIZAJE A ADQUIRIR POR EL ALUMNO

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
- CB8. Que los estudiantes sean capaces de integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o limitada, incluya reflexiones sobre las responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos y juicios
- 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

RESULTADOS DE APRENDIZAJE:

El alumno será capaz de:

- El alumno será capaz de interpretar los modelos de clasificación de datos utilizando las técnicas estudiadas
- El alumno conocerá los métodos principales utilizados en minería de datos y su utilidad en casos reales

BIBLIOGRAFÍA Y RECURSOS DE REFERENCIA GENERALES

BIBLIOGRAFÍA BÁSICA:

- Laura Igual, Santi Seguí (2017): Introduction to Data Science. Springer. ISBN: 9783319500164
- Steven S. Skiena (2017): The Data Science Design Manual. Springer. ISBN: 9783319554433

BIBLIOGRAFÍA COMPLEMENTARIA:

- Kevin P. Murphy (2012): Machine Learning: A Probabilistic Perspective. MIT Press. ISBN: 9780262018029
- Max Kuhn, Kjell Johnson (2013): Applied Predictive Modelling. Springer. ISBN: 9781461468486
- Andreas C. Müller, Sarah Guido (2017): Introduction to Machine Learning with Python: A Guide for Data Scientists. O'Reilly Media. ISBN: 9781449369880

WEBS DE REFERENCIA:

Web / Descripción

[Scikit-learn](https://scikit-learn.org/)(<https://scikit-learn.org/>)

Documentación y guías de la librería Python de aprendizaje automático scikit-learn

PLANIFICACIÓN DEL PROCESO DE ENSEÑANZA-APRENDIZAJE DE LA ASIGNATURA

METODOLOGÍAS:

Se describe a continuación la metodología aplicada

MÉTODO DIDÁCTICO:

El papel del profesor cobra importancia a través de la impartición de clases magistrales en tiempo real por videoconferencia que podrá utilizar para explicar los contenidos teóricos, resolver dudas que se planteen durante la sesión, ofrecer retroalimentación sobre las actividades de evaluación continua o realizar sesiones de tutoría de carácter grupal.

MÉTODO DIALÉCTICO:

Se caracteriza por la participación de los alumnos en las actividades de evaluación continua de debate y la intervención de éstos a través del diálogo y de la discusión crítica (seminarios, grupos de trabajo, etc.). Utilizando este método el alumno adquiere conocimiento mediante la confrontación de opiniones y puntos de vista. El papel del profesor consiste en proponer a través de Open Campus temas referidos a la materia objeto de estudio que son sometidos a debate para, posteriormente, evaluar el grado de comprensión que han alcanzado los alumnos.

MÉTODO HEURÍSTICO:

Este método puede desarrollarse de forma individual o en grupo a través de las actividades de evaluación continua (entregas de trabajos, resolución de ejercicios, presentaciones, etc.). El objetivo es que el alumno asuma un papel activo en el proceso de aprendizaje adquiriendo los conocimientos mediante la experimentación y la resolución de problemas.

CONSIDERACIONES DE LA PLANIFICACIÓN:

Las actividades formativas que se realizan en la asignatura son las siguientes:

Clases teóricas: Actividad dirigida por el profesor que se desarrollará de forma sincrónica en grupo. Para la realización de esta actividad en Open Campus, la UEMC dispone de herramientas de Webconference que permiten una comunicación unidireccional en las que el docente puede desarrollar sesiones en tiempo real con posibilidad de ser grabadas para ser emitidas en diferido.

Actividades prácticas: Actividades supervisadas por el profesor que se desarrollarán fundamentalmente de forma asíncrona, y de forma individual o en grupo:

- ○ Actividades de debate. Se trata de actividades en las que se genera conocimiento mediante la participación de los estudiantes en discusiones alrededor de temas de interés en las distintas asignaturas.
- Entregas de trabajos individuales o en grupos a partir de un enunciado o unas pautas de trabajo que establecerá el profesor.
- Resolución de ejercicios y problemas que el alumno debe realizar a través de Open Campus en un periodo de tiempo determinado. Esta actividad puede ser en formato test de evaluación.

Tutorías: Las tutorías podrán tener un carácter sincrónico o asíncrono y podrán desarrollarse de manera individual o en grupos reducidos.

Están previstas dos sesiones de tutoría por videoconferencia, una al inicio y otra al final del semestre. En la primera se presentará la asignatura y la guía docente y en la segunda, en las semanas previas a la evaluación final, se dedicará a la resolución de dudas de los estudiantes.

Además, el docente utiliza el Tablón, el Foro y el Sistema de correo interno de Open Campus para atender las necesidades y dudas académicas de los estudiantes.

SESIONES EN TIEMPO REAL :

Título	
TU1	Presentación asignatura y Guía docente
CM1	Fundamentos de Machine Learning
CM2	Regularización
CM3	Selección de características
CM4	Árboles de Clasificación y Regresión
CM5	Random Forest
CM6	Boosted Trees
CM7	Máquinas de Vectores Soporte y Métodos kernel
TU2	Resolución de dudas antes de la evaluación

EVALUACIÓN CONVOCATORIA ORDINARIA:

Evaluación continua	60%
Evaluación final	40%

ACTIVIDADES Y SISTEMAS DE EVALUACIÓN :

Tipo Evaluación	Nombre Actividad	% Calif.
Evaluación continua (60 %)	1. Actividad 1 (Entrega individual)	20
	2. Actividad 2 (Entrega individual)	20
	3. Foro de debate grupal a realizar en 2 días (Foro)	10
	4. Test de evaluación (Test de evaluación)	10
Evaluación final (40 %)	1. Prueba de evaluación final (Prueba de evaluación final)	40

CONSIDERACIONES EVALUACIÓN CONVOCATORIA ORDINARIA:

A lo largo de la planificación de la asignatura el alumno realizará **actividades de evaluación continua** que forman parte de la calificación de la asignatura con un peso del 60% sobre la nota final.

Para superar la evaluación continua, el alumno debe obtener una media de igual o superior a 5 entre todas las actividades. En el caso de no superar la evaluación continua, se guardan para la convocatoria extraordinaria las notas de aquellas actividades aprobadas, no pudiendo volver a presentarlas.

El sistema de evaluación de esta asignatura acentúa el desarrollo gradual de competencias y resultados de aprendizaje y, por tanto, se realizará una evaluación continua a través de las distintas actividades de evaluación propuestas. El resultado de la evaluación continua se calcula a partir de las notas obtenidas en cada actividad teniendo en cuenta el porcentaje de representatividad en cada caso.

Todas las actividades deberán entregarse en las fechas previstas para ello, teniendo en cuenta:

- Las actividades de evaluación continua (entrega de trabajos) se desarrollarán según se indica y, para ser evaluadas, los trabajos deberán ser entregados en la forma y fecha prevista y con la extensión máxima señalada. No se evaluarán trabajos entregados posteriormente a esta fecha o que no cumplan con los criterios establecidos por el profesor.
- La no entrega de una actividad de evaluación continua en forma y plazo se calificará con un 0 y así computarán en el cálculo de la nota de evaluación continua y final de la asignatura.
- Cualquier tipo de copia o plagio por mínimo que sea, así como un uso inapropiado de herramientas de inteligencia artificial, supondrá una calificación de 0 en la actividad correspondiente. Esta actuación podría suponer la apertura de un expediente disciplinario.
- Las actividades de evaluación continua (tipo test) se desarrollarán con anterioridad a la realización de las pruebas de evaluación final de la asignatura
- La participación en los foros y debates vinculados a una actividad de evaluación continua se evaluará de forma individual o colectiva según se haya requerido la participación: el valor de las aportaciones, el número de aportaciones y respuestas en debate a las opiniones de los compañeros. Cualquier comentario aportado en el foro que suponga una falta de respeto a las opiniones de compañeros supondrá el suspenso de la evaluación continua de la asignatura.

Los alumnos accederán a través de OpenCampus a las calificaciones de las actividades de evaluación continua en un plazo no superior a 15 días lectivos desde su fecha de entrega, excepto causas de fuerza mayor en cuyo caso se informará al alumno a través del Tablón.

La evaluación continua se completará con una **evaluación final**, que se realizará al finalizar el periodo lectivo de cada asignatura. Los exámenes serán eminentemente prácticas, de manera que, los alumnos podrán disponer de los apuntes y consultarlos, (solo en formato digital) durante la realización de la prueba.

Para resolver el examen, los alumnos deberán descargar el enunciado de la prueba y una vez cumplimentado, subirlo en el espacio correspondiente del campus virtual

La prueba supondrá un 40% de la calificación sobre la nota final de la asignatura.

- El alumno tendrá la posibilidad, siempre dentro de los tres días siguientes a la publicación de las notas, a renunciar a su calificación, y presentarse en la siguiente convocatoria
- El alumno tendrá hasta 3 días después de la calificación para solicitar al docente más información sobre su calificación por el correo de la plataforma. DOCENTE
- Cualquier tipo de irregularidad o fraude en la realización de una prueba, así como un uso inapropiado de herramientas de inteligencia artificial, supondrá una calificación de 0 en la prueba/convocatoria correspondiente. Esta actuación podría suponer la apertura de un expediente disciplinario.
- El aplazamiento concedido por la Universidad para la realización de una evaluación final se regirá por lo establecido en el Manual de "Directrices y plazos para la tramitación de una solicitud"

La **nota global** de la asignatura se obtiene ponderando la calificación de la evaluación continua y de la evaluación final según los siguientes porcentajes, y debiendo tener aprobadas ambas partes, continua y final, para superar la asignatura.

Si un alumno no se presenta a la prueba de evaluación final, su calificación en la convocatoria será de "No presentado", con independencia de que haya realizado alguna actividad de evaluación continua.

De igual modo si el alumno no entrega ninguna actividad de evaluación continua, obtendrá la calificación de "No presentado", con independencia de que haya aprobado la prueba de evaluación final, en cuyo caso, se le guardaría su calificación para la convocatoria extraordinaria.

EVALUACIÓN CONVOCATORIA EXTRAORDINARIA:

Evaluación continua	60%
Evaluación final	40%

ACTIVIDADES Y SISTEMAS DE EVALUACIÓN :

Tipo Evaluación	Nombre Actividad	% Calif.
Evaluación continua (60 %)	1. Actividad 1 (Entrega individual)	20
	2. Actividad 2 (Entrega individual)	20
	3. Actividad 3 (Entrega individual)	10
	4. Test de evaluación (Test de evaluación)	10
Evaluación final (40 %)	1. Prueba de evaluación final (Prueba de evaluación final)	40

CONSIDERACIONES EVALUACIÓN CONVOCATORIA EXTRAORDINARIA:

Los estudiantes que no hayan superado la asignatura en la convocatoria ordinaria, porque hayan suspendido la evaluación continua o la prueba de evaluación final, podrán presentarse a las pruebas establecidas por el profesor en la convocatoria extraordinaria.

Para la convocatoria extraordinaria se guardan las calificaciones de las actividades de evaluación continua y prueba de evaluación final, superadas por el estudiante (nota superior o igual a 5), no permitiéndose volver a realizarlas.

- En convocatoria extraordinaria, el alumno solo podrá entregar las actividades de evaluación continua no superadas, guardándose la calificación de las aprobadas.
- El alumno tendrá hasta 3 días después de la calificación para solicitar al docente más información sobre su calificación por el correo de la plataforma.

- Cualquier tipo de irregularidad o fraude en la realización de una prueba, supondrá una calificación de 0 en la prueba/convocatoria correspondiente.
- El aplazamiento concedido por la Universidad para la realización de una evaluación final se registrará por lo establecido en el Manual de "Directrices y plazos para la tramitación de una solicitud".

En la convocatoria extraordinaria, la **nota global** de la asignatura se obtiene ponderando la calificación de la evaluación continua y de la evaluación final, de la misma forma que en la convocatoria ordinaria.

Al igual que en la convocatoria ordinaria, en la convocatoria extraordinaria es necesario superar tanto la evaluación continua como la evaluación final para aprobar la asignatura.

Si un alumno no se presenta a la prueba de evaluación final, su calificación en la convocatoria será de "No presentado", con independencia de que haya realizado alguna actividad de evaluación continua.

SISTEMAS DE EVALUACIÓN:

SISTEMA DE EVALUACIÓN	PORCENTAJE (%)
Escalas de actitudes	10%
Pruebas de ejecución de tareas reales y/o simuladas	20%
Pruebas de respuesta corta	20%
Pruebas de respuesta larga, de desarrollo	20%
Pruebas objetivas	10%
Trabajos y proyectos	20%