

CENTRO UNIVERSITARIO DE TECNOLOGÍA Y ARTE DIGITAL



PLANIFICACIÓN DE LA DOCENCIA UNIVERSITARIA

GUÍA DOCENTE

Aprendizaje Automático II

1. DATOS DE IDENTIFICACIÓN DE LA ASIGNATURA.

Título:	Grado en Ingeniería del Software
Facultad:	Centro Universitario de Tecnología y Arte Digital (U-TAD)
Materia:	Ingeniería de datos
Denominación de la asignatura:	Aprendizaje Automático II
Curso:	4º
Cuatrimestre:	2
Carácter:	Obligatoria de mención
Créditos ECTS:	6
Modalidad/es de enseñanza:	Híbrida Presencial
Idioma:	Castellano
Profesor/a:	Pedro Concejero Cerezo
E-mail:	pedro.concejero@u-tad.com
Teléfono:	

2. DESCRIPCIÓN DE LA ASIGNATURA

2.1 Descripción de la materia

Los contenidos de la materia permiten a los alumnos comprender el flujo de búsqueda, ingesta, almacenamiento, procesamiento y análisis de información de datos y aproxima a los alumnos a las técnicas y tecnologías necesarias para la gestión de grandes cantidades de datos.

2.2 Descripción de la asignatura

El objetivo es conocer los modelos y algoritmos más recientes para la solución de problemas de carácter empresarial o académico **siempre basados en datos**. La asignatura también revisará las herramientas y entornos modernos para el análisis de datos.

El enfoque será:

- Totalmente aplicado, enfocado en la utilización de los algoritmos, librerías o entornos más que en el desarrollo matemático de los mismos -aunque por supuesto revisaremos todo lo necesario para comprender adecuadamente su campo de aplicación y requisitos.
- Basado en el dato, en datos lo más reales posibles.
- Enfocado a casos y ejemplos prácticos, desde el pre-proceso de datos hasta la evaluación y comparación de los modelos.
- Lo más rico posible en herramientas y entornos. Como mínimo, R y python (scikit-learn), pero también otros entornos recientes que sean de interés.

3. COMPETENCIAS Y RESULTADOS DE APRENDIZAJE

Competencias Básicas y Generales
<p>CB1: 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.</p> <p>CB2: 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.</p> <p>CB3: 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.</p> <p>CB4: Que los estudiantes puedan transmitir información, ideas, problemas y soluciones a un público tanto especializado como no especializado.</p> <p>CB5: Que los estudiantes hayan desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía</p> <p>CG1 - Capacidad para entender, planificar y resolver problemas a través del desarrollo de soluciones informáticas.</p> <p>CG3 - Conocimiento de los fundamentos científicos aplicables a la resolución de problemas informáticos</p> <p>CG4 - Capacidad para simplificar y optimizar los sistemas informáticos atendiendo a la comprensión de su complejidad</p> <p>CG9 - Capacidad para aprender, modificar y producir nuevas tecnologías informáticas</p> <p>CG10 - Uso de técnicas creativas para la realización de proyectos informáticos</p> <p>CG11 - Capacidad de buscar, analizar y gestionar la información para poder extraer conocimiento de la misma</p>
Competencias Específicas
<p>CE3 - Conocimiento del álgebra relacional y realización de consultas en lenguajes procedurales para el diseño de esquemas de bases de datos normalizados basados en modelos de entidad-relación</p> <p>CE10 - Capacidad para manejar un gestor de versiones de código y generar la documentación de una aplicación de forma automática.</p>
Competencias Transversales
<p>CT1 - Conocimiento de la definición, el alcance y la puesta en práctica de los fundamentos de las metodologías de gestión de proyectos de desarrollo tecnológico</p> <p>CT2 - Conocimiento de los principales agentes del sector y del ciclo de vida completo de un proyecto de desarrollo y comercialización de contenidos digitales</p> <p>CT4 - Capacidad de actualización del conocimiento adquirido en el manejo de herramientas y tecnologías digitales en función del estado actual del sector y de las tecnologías empleadas</p> <p>CT5 - Desarrollo de las habilidades necesarias para el emprendimiento digital.</p>

4. CONTENIDOS

4.1. Temario de la asignatura

Tools 1. R

T.1.1. rattle

T.1.2. advanced R

T.1.3. creating R libraries

T.1.4. The Tidyverse approach to data science (recap)

Tools 2. python

T.2.1. scikit-learn

T.2.2. scikit-learn and pyspark

Tools 3. Automatic ML platforms

T.3.1. H2O (<https://www.h2o.ai/>)

T.3.2. AutoML Frameworks in R & Python (<https://www.r-bloggers.com/2020/04/automl-frameworks-in-r-python/>)

Models 1. Performance Measures for evaluating models

M.1.1. ROC curves and related measures (binary classification)

M.1.2. Cost curves

M.1.3. Multicategorical classification

M.1.4. Regression-type evaluation measures

Models 2. Screening

Models 3. Modern classification

M.3.1. SVMs (linear and non-linear kernels)

M.3.2. Random Forests

M.3.3. Gradient Boosting

Models 4. Time to event. Survival analysis

Models 5. Dynamic models. Time Series. ARIMA and related models

Models 6. Regression-type models in 2021.

Models 7. Collaborative Filtering. Recommenders

Models 8. Non-supervised: autoencoders

Models 9. Ensembles

4.2. Desarrollo temporal

UNIDADES DIDÁCTICAS / TEMAS	PERÍODO TEMPORAL
T.1.1. rattle M.1.1. ROC M.2 Screening Tarea 1 (entrega 24 febrero)	10-25 febrero
M.1.2 Cost Curves M.3 SVM, RF, GB for binary classification	3-4 marzo
M.1.3 Multi-category classification performance measures M.3 Models for multcategory classification (same as those for binary) Tarea 2 (entrega 11 marzo)	10-11 marzo
M.1.4 Regression evaluation measures M.4 Survival analysis (time to event)	17-18 marzo
M.6 Regression Models Tarea 3 (entrega 25 marzo)	24-25 marzo
Semana Santa	29 marzo – 4 abril
M.5 Time series ARIMA and related models	7-8 abril
T.1.2 Advanced R T.1.3 Creating R libraries T.1.4 Tidyverse (recap)	14–15 abril
M.7 Collaborative Filtering and recommenders Tarea 4 (entrega 22 abril)	21-22 abril
T.2 python scikit-learn (& pyspark)	28-29 abril
T.3 Automatic ML platforms Tarea 5 (entrega 5 de mayo)	5 -6 mayo
M.9 Ensembles	12-13 mayo
M.8 Non-supervised: autoencoders Repaso y foco en trabajo final Tarea 6 (entrega 20 mayo)	19-20 mayo
Entrega trabajo final – Fecha límite	Lunes 24 de mayo

5. ACTIVIDADES FORMATIVAS Y MODALIDADES DE ENSEÑANZAS

5.1. Modalidades de enseñanza

La asignatura se desarrollará a través de los siguientes métodos y técnicas generales, que se aplicarán diferencialmente según las características propias de la asignatura:

- Método expositivo/Lección magistral: el profesor desarrollará, mediante clases magistrales y dinámicas los contenidos recogidos en el temario.
- Estudio de casos: análisis de casos reales relacionados con la asignatura.
- Resolución de ejercicios y problemas: los estudiantes desarrollarán las soluciones adecuadas aplicando procedimientos de transformación de la información disponible y la interpretación de los resultados.
- Aprendizaje basado en problemas: utilización de problemas como punto de partida para la adquisición de conocimientos nuevos.
- Aprendizaje orientado a proyectos: se pide a los alumnos que, en pequeños grupos, planifiquen, creen y evalúen un proyecto que responda a las necesidades planteadas en una determinada situación.
- Aprendizaje cooperativo: Los estudiantes trabajan en grupo para realizar las tareas de manera colectiva.

Actividades formativas

Actividad Formativa	Horas	Presencialidad
AF1 Clases teóricas / Expositivas	45	100%
AF2 Clases Prácticas	36	100%
AF3 Tutorías	9	50%
AF4 Estudio independiente y trabajo autónomo del alumno	57,5	0%
AF5 Elaboración de trabajos (en grupo o individuales)	28,5	0%
AF6: Actividades de Evaluación	9	100%

6. SISTEMA DE EVALUACIÓN

ACTIVIDAD DE EVALUACIÓN	VALORACIÓN MÍNIMA RESPECTO A LA CALIFICACIÓN FINAL (%)	VALORACIÓN MÁXIMA RESPECTO A LA CALIFICACIÓN FINAL (%)
SE1 Evaluación de la participación en clase, en prácticas o en proyectos de la asignatura	10%	30%
SE2 Evaluación de trabajos, proyectos, informes, memorias	40%	80%
SE3 Prueba Objetiva	10%	60%

6.1. Criterios de calificación

ACTIVIDAD DE EVALUACIÓN	VALORACIÓN RESPECTO A LA CALIFICACIÓN FINAL (%)
Evaluación de la participación en clase, en prácticas o en proyectos de la asignatura	10%
Evaluación de trabajos, proyectos, informes, memorias	45%
Prueba Objetiva	45%

Consideraciones generales acerca de la evaluación:

- La calificación estará basada en dos tipos de entregas o actividades:
 - Tareas de evaluación continua, quincenales, hasta un máximo de 7. Todas ellas serán calificadas.
 - Trabajo final que se presentará con fecha límite improrrogable por requisito UTAD- el día 24-5-2021. La composición y alcance de este trabajo será objeto de sesiones específicas en el desarrollo de las clases.
- Para aprobar la asignatura en la convocatoria ordinaria, es imprescindible que tanto la nota de las actividades a entregar como la del trabajo final sea como mínimo de 5.0 (sobre 10).
- En caso de no conseguir el aprobado en la convocatoria ordinaria, el alumno podrá presentarse a la convocatoria extraordinaria de julio, donde realizará un nuevo trabajo final que representará el 100% de su calificación en dicha convocatoria.

7. BIBLIOGRAFÍA / WEBGRAFÍA

Bibliografía básica

R manuals, <http://cran.r-project.org/manuals.html>

Trevor Hastie, Robert Tibshirani, Jerome Friedman: The Elements of Statistical Learning. Data Mining, Inference, and Prediction. 2nd. Edition. Springer Series in Statistics

<https://web.stanford.edu/~hastie/ElemStatLearn/>

(por favor asegúrate que se descargue la versión “corrected 12th printing Jan 2017”)

Andriy Burkov, ‘The Hundred-Page Machine Learning Book’ 2019. ISBN-13:978-1999579500

Advanced R, 2nd edition. Hadley Wickham. <https://adv-r.hadley.nz/>

The tidyverse approach to data science

<https://tidyverse.tidyverse.org/articles/paper.html>

scikit-learn Tutorials <https://scikit-learn.org/stable/tutorial/index.html>

8.- MATERIAL, SOFTWARE Y HERRAMIENTAS NECESARIAS

TIPOLOGÍA DEL AULA:

Sala virtual del curso en Blackboard

MATERIALES DEL ALUMNO:

Ordenador personal, webcam y micrófono

SOFTWARE:

R 4.0 <https://cran.r-project.org/>

RStudio <https://rstudio.com/products/rstudio/download/>

rattle <https://rattle.togaware.com/>

<http://cran.r-project.org/web/packages/rattle/index.html>