



UNIVERSIDAD
POLITÉCNICA
DE MADRID

PROCESO DE
COORDINACIÓN DE LAS
ENSEÑANZAS PR/CL/001



E.T.S. de Ingenieros
Informaticos

ANX-PR/CL/001-01

GUÍA DE APRENDIZAJE

ASIGNATURA

103000750 - Inteligencia Artificial Explicable

PLAN DE ESTUDIOS

10AJ - Master Universitario En Inteligencia Artificial

CURSO ACADÉMICO Y SEMESTRE

2024/25 - Primer semestre

Índice

Guía de Aprendizaje

1. Datos descriptivos.....	1
2. Profesorado.....	1
3. Conocimientos previos recomendados.....	2
4. Competencias y resultados de aprendizaje.....	2
5. Descripción de la asignatura y temario.....	3
6. Cronograma.....	5
7. Actividades y criterios de evaluación.....	7
8. Recursos didácticos.....	9
9. Otra información.....	10

1. Datos descriptivos

1.1. Datos de la asignatura

Nombre de la asignatura	103000750 - Inteligencia Artificial Explicable
No de créditos	5 ECTS
Carácter	Optativa
Curso	Primer curso
Semestre	Primer semestre
Período de impartición	Septiembre-Enero
Idioma de impartición	Castellano
Titulación	10AJ - Master Universitario en Inteligencia Artificial
Centro responsable de la titulación	10 - Escuela Tecnica Superior De Ingenieros Informaticos
Curso académico	2024-25

2. Profesorado

2.1. Profesorado implicado en la docencia

Nombre	Despacho	Correo electrónico	Horario de tutorías *
Bojan Mihaljevic	3205	b.mihaljevic@upm.es	Sin horario. Sin horario. http://www.dia.fi.upm.es/es/bmihaljevic
Esteban Garcia Cuesta (Coordinador/a)	2201	esteban.garcia@upm.es	Sin horario. Sin horario. http://www.dia.fi.upm.es/es/estebangarcia

* Las horas de tutoría son orientativas y pueden sufrir modificaciones. Se deberá confirmar los horarios de tutorías con el profesorado.

3. Conocimientos previos recomendados

3.1. Asignaturas previas que se recomienda haber cursado

El plan de estudios Master Universitario en Inteligencia Artificial no tiene definidas asignaturas previas recomendadas para esta asignatura.

3.2. Otros conocimientos previos recomendados para cursar la asignatura

- La asignatura se imparte en español pero debido a que existe mucha literatura en inglés (la cual es de apoyo para la impartición y seguimiento de la asignatura) se recomienda tener conocimiento del idioma inglés para lectura.
- Nociones del lenguaje de programación Python
- Se recomienda tener conocimientos básicos de inteligencia artificial y aprendizaje automático

4. Competencias y resultados de aprendizaje

4.1. Competencias

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.

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

CEIA4 - Capacidad de interpretar los modelos de clasificación supervisada y no supervisada obtenidos al aplicar las técnicas de Aprendizaje Automático para un conjunto de datos.

CG11 - Adquirir conocimientos científicos avanzados del campo de la informática que le permitan generar nuevas

ideas dentro de una línea de investigación.

4.2. Resultados del aprendizaje

RA109 - Ser capaz de aportar nuevas ideas y conocimiento tanto a nivel metodológico, como de aplicación de la explicabilidad en el campo de la Inteligencia Artificial

RA112 - Conocer las técnicas de explicabilidad de la Inteligencia Artificial y sus aplicaciones

RA110 - Ser capaz de dotar con explicabilidad a los modelos de aprendizaje automático

RA113 - Ser capaz de conocer las principales ventajas y desventajas de las técnicas de explicabilidad y tener las bases para entender futuros avances

5. Descripción de la asignatura y temario

5.1. Descripción de la asignatura

A pesar del alto rendimiento de los algoritmos de Inteligencia Artificial (IA) muchos de ellos son incomprensibles para las personas. La Inteligencia Artificial Explicable (IAX) trata de desarrollar sistemas inteligentes artificiales que sean inherentemente explicables, así como técnicas que permitan la explicación de modelos de aprendizaje automático considerados del tipo "caja negra". El concepto de "caja negra" se utiliza para referirse al hecho de que no se puede explicar el motivo de una decisión final dado un modelo aprendido. La Inteligencia Artificial Explicable genera elementos que explican la decisión de un modelo aprendido proporcionando un conocimiento que mejora su inteligibilidad y asegura su validez. Entre las aplicaciones de interés pueden citarse el análisis clínico, la bioinformática, la navegación autónoma, o sistemas de ayuda al descubrimiento de conocimiento. En esta asignatura se ven distintos modelos y métodos que permiten interpretar y dotar con explicabilidad a un sistema inteligente artificial. El objetivo es conocer estos métodos desde una perspectiva global, sus usos y limitaciones, y algunas métricas que permiten su validación.

5.2. Temario de la asignatura

1. Introducción a la Inteligencia Artificial eXplicable: conceptos, ideas y principios
2. Modelos interpretables
3. Técnicas de explicabilidad post-hoc
4. Técnicas de explicabilidad basadas en ejemplos
5. Métricas y evaluación de la explicabilidad
6. Modelos basados en conceptos y prototipos
7. Tendencias, aplicaciones, e investigación en la actualidad

6. Cronograma

6.1. Cronograma de la asignatura *

Sem	Actividad tipo 1	Actividad tipo 2	Tele-enseñanza	Actividades de evaluación
1	Clases tema 1 Duración: 02:00 LM: Actividad del tipo Lección Magistral			
2	Clases tema 2 Duración: 02:00 LM: Actividad del tipo Lección Magistral			
3	Clases tema 2 Duración: 02:00 LM: Actividad del tipo Lección Magistral			
4	Clases tema 2 Duración: 02:00 LM: Actividad del tipo Lección Magistral			
5				Entrega de actividad temas 1-2 Evaluación progresiva de la actividad con presentación PG: Técnica del tipo Presentación en Grupo Evaluación Progresiva Presencial Duración: 02:00
6	Clases tema 3 Duración: 02:00 LM: Actividad del tipo Lección Magistral			
7	Clases tema 3 Duración: 02:00 LM: Actividad del tipo Lección Magistral			
8	Clases tema 3 Duración: 02:00 LM: Actividad del tipo Lección Magistral Actividades Transversales Duración: 02:00 OT: Otras actividades formativas / Evaluación			
9	Clases tema 3 Duración: 02:00 LM: Actividad del tipo Lección Magistral			
10	Clases tema 4 Duración: 02:00 LM: Actividad del tipo Lección Magistral Actividades Transversales Duración: 02:00 OT: Otras actividades formativas / Evaluación			

11	<p>Clases tema 4 Duración: 02:00 LM: Actividad del tipo Lección Magistral</p>			
12	<p>Clases tema 4 Duración: 02:00 LM: Actividad del tipo Lección Magistral</p> <p>Actividades Transversales Duración: 02:00 OT: Otras actividades formativas / Evaluación</p>			
13	<p>Clases tema 5 Duración: 02:00 LM: Actividad del tipo Lección Magistral</p>			
14	<p>Clases tema 6 Duración: 02:00 LM: Actividad del tipo Lección Magistral</p> <p>Actividades Transversales Duración: 02:00 OT: Otras actividades formativas / Evaluación</p>			
15	<p>Clases tema 7 Duración: 02:00 LM: Actividad del tipo Lección Magistral</p>			
16				<p>Entrega de actividad temas 3-6 Evaluación progresiva de la actividad con presentación PG: Técnica del tipo Presentación en Grupo Evaluación Progresiva Presencial Duración: 01:00</p> <p>Examen teórico de evaluación progresiva, parcial temas 1-6 EX: Técnica del tipo Examen Escrito Evaluación Progresiva Presencial Duración: 01:00</p>
17				<p>Examen teórico-práctico de evaluación progresiva, parcial temas 1-2 EX: Técnica del tipo Examen Escrito Evaluación Global Presencial Duración: 01:00</p> <p>Examen teórico-práctico de evaluación progresiva, parcial temas 3-6 EX: Técnica del tipo Examen Escrito Evaluación Global Presencial Duración: 01:00</p>

Para el cálculo de los valores totales, se estima que por cada crédito ECTS el alumno dedicará dependiendo del plan de estudios, entre 26 y 27 horas de trabajo presencial y no presencial.

7. Actividades y criterios de evaluación

7.1. Actividades de evaluación de la asignatura

7.1.1. Evaluación (progresiva)

Sem.	Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
5	Entrega de actividad temas 1-2 Evaluación progresiva de la actividad con presentación	PG: Técnica del tipo Presentación en Grupo	Presencial	02:00	20%	4 / 10	CB10 CEIA4
16	Entrega de actividad temas 3-6 Evaluación progresiva de la actividad con presentación	PG: Técnica del tipo Presentación en Grupo	Presencial	01:00	47%	4 / 10	CGI1 CB6
16	Examen teórico de evaluación progresiva, parcial temas 1-6	EX: Técnica del tipo Examen Escrito	Presencial	01:00	33%	4 / 10	CGI1 CB10 CB7 CB6

7.1.2. Prueba evaluación global

Sem	Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
17	Examen teórico-práctico de evaluación progresiva, parcial temas 1-2	EX: Técnica del tipo Examen Escrito	Presencial	01:00	30%	4 / 10	CEIA4 CB6 CB10 CB7
17	Examen teórico-práctico de evaluación progresiva, parcial temas 3-6	EX: Técnica del tipo Examen Escrito	Presencial	01:00	70%	4 / 10	CB7 CGI1 CB6 CB10

7.1.3. Evaluación convocatoria extraordinaria

Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
-------------	-----------	------	----------	-----------------	-------------	------------------------

Examen teórico-práctico de evaluación extraordinaria	EX: Técnica del tipo Examen Escrito	Presencial	02:00	100%	5 / 10	CEIA4 CB6 CB10 CB7 CGI1
--	-------------------------------------	------------	-------	------	--------	-------------------------------------

7.2. Criterios de evaluación

La calificación final vendrá dada a partir de la evaluación de dos partes ponderadas de acuerdo a las tabla de evaluación detalladas en el apartado anterior. Para poder ponderar la nota de las dos partes es necesario que el estudiante obtenga una nota mínima de 4 sobre 10 en cada parte.

Las calificaciones de 5 o superiores en exámenes y/o proyectos prácticos parciales se conservan para la evaluación global y extraordinaria. Solo se puede optar a una segunda o tercera evaluación de exámenes parciales/proyectos prácticos si estos han recibido calificaciones inferiores a 5 o no se han presentado.

Los enunciados de los proyectos prácticos podrán variar en las distintas evaluaciones (progresiva, global, y extraordinaria).

Los derechos y deberes de los estudiantes universitarios están desarrollados en los Estatutos de la Universidad Politécnica de Madrid (BOCM de 15 de noviembre de 2010) y en el Estatuto del Estudiante Universitario (RD 1791/2010 de 30 de diciembre). El artículo 124 (a) de los EUPM fija como deber del estudiante ..."seguir con responsabilidad y aprovechamiento el proceso de formación, adquisición de conocimientos, y aprendizaje correspondiente a su condición de universitario"... y el artículo 13 del Estatuto del Estudiante Universitario, en el punto (d) especifica también como deber del estudiante universitario "abstenerse de la utilización o cooperación en procedimientos fraudulentos en las pruebas de evaluación, en los trabajos que se realicen o en documentos oficiales de la universidad". En el caso de que en el desarrollo de las pruebas de evaluación se aprecie el incumplimiento de los deberes como estudiante universitario, el coordinador de la asignatura podrá ponerlo en conocimiento del Director o Decano del Centro, que de acuerdo con lo establecido en el artículo 74 (n) de los Estatutos de la UPM tiene competencias para "Proponer la iniciación del procedimiento disciplinario a cualquier miembro de la Escuela o Facultad, por propia iniciativa o a instancia de la Comisión de Gobierno" al Rector, en los términos previstos en los estatutos y normas de aplicación.

8. Recursos didácticos

8.1. Recursos didácticos de la asignatura

Nombre	Tipo	Observaciones
1. Interpretable Machine Learning: A Guide For Making Black Box Models Explainable.	Bibliografía	Molnar, C., 2020. Interpretable Machine Learning: A Guide For Making Black Box Models Explainable. https://christophm.github.io/interpretable-ml-book/
2. Concepts, taxonomies, opportunities and challenges toward responsible AI.	Bibliografía	Arrieta, A. B. et.al. Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58:82-115. arXiv:1910.10045
3. An introduction to Machine Learning Interpretability	Bibliografía	Hall, P. and Gill, N. An Introduction to Machine Learning Interpretability. Ed. O'Reilly (2019) https://h2o.ai/content/dam/h2o/en/marketing/documents/2019/08/An-Introduction-to-Machine-Learning-Interpretability-Second-Edition.pdf
4. A unified approach to interpreting model predictions	Bibliografía	Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. arXiv:1705.07874
5. Principles and Practice of Explainable Machine Learning	Bibliografía	Belle, Vaishak & Papantonis, Ioannis. (2020). Principles and Practice of Explainable Machine Learning. arXiv:2009.11698

9. Otra información

9.1. Otra información sobre la asignatura

La asignatura se apoya en la herramienta Moodle para proporcionar información y documentación a los alumnos, así como para la asignación de enunciados y entregas de las prácticas y la comunicación de las calificaciones de los alumnos.