



UNIVERSIDAD
POLITÉCNICA
DE MADRID

PROCESO DE
COORDINACIÓN DE LAS
ENSEÑANZAS PR/CL/001



E.T.S. de Ingeniería de
Sistemas Informáticos

ANX-PR/CL/001-01

GUÍA DE APRENDIZAJE

ASIGNATURA

613000123 - Procesamiento Del Lenguaje Natural Mediante Deep Learning

PLAN DE ESTUDIOS

61AH - Máster Universitario En Aprendizaje Automático Y Datos Masivos

CURSO ACADÉMICO Y SEMESTRE

2023/24 - 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.....	4
6. Cronograma.....	6
7. Actividades y criterios de evaluación.....	8
8. Recursos didácticos.....	10

1. Datos descriptivos

1.1. Datos de la asignatura

Nombre de la asignatura	613000123 - Procesamiento del Lenguaje Natural Mediante Deep Learning
No de créditos	3 ECTS
Carácter	Obligatoria
Curso	Primer curso
Semestre	Primer semestre
Período de impartición	Septiembre-Enero
Idioma de impartición	Castellano
Titulación	61AH - Máster Universitario en Aprendizaje Automático y Datos Masivos
Centro responsable de la titulación	61 - Escuela Técnica Superior De Ingeniería De Sistemas Informáticos
Curso académico	2023-24

2. Profesorado

2.1. Profesorado implicado en la docencia

Nombre	Despacho	Correo electrónico	Horario de tutorías *
Alejandro Martin Garcia (Coordinador/a)	1209	alejandro.martin@upm.es	Sin horario. Solicitar por correo
Javier Huertas Tato	1220	javier.huertas.tato@upm.es	Sin horario. Solicitar por correo

* 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 Máster Universitario en Aprendizaje Automático y Datos Masivos no tiene definidas asignaturas previas recomendadas para esta asignatura.

3.2. Otros conocimientos previos recomendados para cursar la asignatura

- Programación en Python
- Machine Learning

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

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

CE03 - Diseñar y aplicar técnicas avanzadas de aprendizaje profundo para la resolución de problemas relacionados con el análisis y tratamiento de datos masivos.

CG1 - Capacidad para aplicar el método científico y saber organizar y planificar experimentos con rigor metodológico en el ámbito del aprendizaje automático y los datos masivos

CG2 - Participar en la aplicación de mecanismos de descripción, cuantificación, análisis, interpretación y evaluación de resultados experimentales del ámbito de los datos masivos y el aprendizaje automático

CG3 - Capacidad para reunir e interpretar datos masivos relevantes para emitir juicios que incluyan una reflexión sobre temas importantes de índole científico, social o ético en el ámbito del aprendizaje automático y los datos masivos

CG4 - Capacidad de aplicar iniciativa, integración, colaboración y potenciación de la discusión crítica en el ámbito del trabajo en equipo dentro del ámbito del aprendizaje automático y datos masivos

CG5 - Participar en la transmisión de la información generada, las ideas, los problemas y las soluciones de forma oral y escrita para un público tanto especializado como no especializado

CT1 - Creatividad

CT2 - Organización y planificación

CT3 - Gestión de la información

CT4 - Liderazgo de equipos

CT5 - Trabajo en contextos internacionales

K01 - El alumno clasifica y explica los algoritmos y técnicas de aprendizaje automático, tanto supervisado como no supervisado

K02 - El alumno explica los modelos de aprendizaje profundo y generativos

S03 - El alumno experimenta con modelos de aprendizaje automático para la obtención de conocimiento a partir de conjuntos de datos masivos

S04 - El alumno crea nuevos modelos de aprendizaje automático y ejecuta experimentos para demostrar su viabilidad y mejora de rendimiento con respecto al estado del arte

S05 - El alumno compone y ejecuta el flujo de trabajo necesario para la resolución de un problema de aprendizaje automático

4.2. Resultados del aprendizaje

RA2 - Manejar y entrenar modelos de representación semántica en un espacio vectorial

RA3 - Utilizar modelos de temas para extracción de información

RA4 - Demostrar comprensión sobre los sistemas de aprendizaje profundo, así como las arquitecturas principales usadas en el PLN

RA5 - Demostrar conocimiento sobre los últimos avances del aprendizaje profundo en PLN.

5. Descripción de la asignatura y temario

5.1. Descripción de la asignatura

En esta asignatura, los estudiantes adquieren conocimientos teóricos y prácticos para comprender y abordar los desafíos del PLN mediante el uso de técnicas de aprendizaje profundo. La asignatura se centrará principalmente en el uso de modelos avanzados de PLN, como son las arquitecturas Transformer, que han demostrado un gran éxito en una variedad de tareas de procesamiento del lenguaje natural. Durante el curso, se analizarán distintos aspectos relacionados con el procesamiento del lenguaje natural utilizando modelos Transformer. Se explorarán metodologías y enfoques de PLN, donde los estudiantes aprenderán sobre las diferentes etapas del proceso de procesamiento del lenguaje natural, incluyendo la tokenización, la normalización, la segmentación y la extracción de características.

Se profundizará en el uso de redes neuronales para el procesamiento del lenguaje natural, donde se estudiarán las diferentes arquitecturas neuronales utilizadas en PLN y, en particular, los modelos Transformer. Los estudiantes comprenderán la importancia de las arquitecturas de redes neuronales en el procesamiento del lenguaje natural y cómo aprovechar su capacidad para capturar patrones y relaciones complejas en los datos lingüísticos. Uno de los aspectos clave de los modelos Transformer que se abordarán en la asignatura es el mecanismo de atención, que permite que el modelo se enfoque en partes relevantes de la entrada y capture las relaciones entre las palabras o elementos del lenguaje. Los estudiantes aprenderán cómo funciona el mecanismo de atención y cómo se aplica en los modelos Transformer.

Además, se estudiarán las arquitecturas encoder y decoder de los modelos Transformer, que se utilizan en tareas como la traducción automática y la generación de texto. Los estudiantes comprenderán cómo se estructuran estas arquitecturas y cómo se entrenan para producir resultados de alta calidad en diferentes tareas de procesamiento del lenguaje natural. La asignatura también abordará el tema del transfer learning en el contexto del PLN. Los

estudiantes aprenderán cómo aprovechar los modelos pre-entrenados de PLN, como BERT y RoBERTa, para mejorar el rendimiento en tareas específicas. Se explorarán técnicas de ajuste fino y adaptación de modelos pre-entrenados a dominios específicos.

Además de las tareas específicas de PLN, se estudiará el aprendizaje auto-supervisado, una técnica que utiliza grandes cantidades de datos no etiquetados para entrenar modelos de PLN. Los estudiantes aprenderán cómo los modelos auto-supervisados pueden aprender representaciones lingüísticas de alta calidad y cómo aplicar estos enfoques en la resolución de problemas de PLN.

5.2. Temario de la asignatura

1. Introducción al Procesamiento del Lenguaje Natural

- 1.1. PLN y sus aplicaciones
- 1.2. Historia y evolución de PLN
- 1.3. PLN como subcampo de la Inteligencia Artificial
- 1.4. Consideraciones éticas

2. Metodologías de PLN

- 2.1. Librerías
- 2.2. Preprocesado
- 2.3. Métricas y evaluación

3. Introducción a los modelos de lenguaje

- 3.1. Técnicas basadas en características
- 3.2. Redes neuronales en PLN

4. Modelos de lenguaje

- 4.1. El mecanismo de atención
- 4.2. Encoders y decoders para PLN
- 4.3. Transfer learning
- 4.4. Tareas de PLN
- 4.5. Aprendizaje auto-supervisado en PLN

6. Cronograma

6.1. Cronograma de la asignatura *

Sem	Actividad en aula	Actividad en laboratorio	Tele-enseñanza	Actividades de evaluación
1				
2				
3				
4				
5				
6	Tema 1 - Introducción Duración: 03:00 LM: Actividad del tipo Lección Magistral Tema 2 - Metodologías Duración: 03:00 LM: Actividad del tipo Lección Magistral			
7	Tema 3 - Introducción a los modelos de lenguaje Duración: 03:00 LM: Actividad del tipo Lección Magistral Tema 3 - Introducción a los modelos de lenguaje Duración: 03:00 PL: Actividad del tipo Prácticas de Laboratorio			
8	Tema 4 - Modelos de lenguaje Duración: 03:00 LM: Actividad del tipo Lección Magistral Tema 4 - Modelos de lenguaje Duración: 03:00 PL: Actividad del tipo Prácticas de Laboratorio			
9	Tema 4 - Modelos de lenguaje Duración: 03:00 LM: Actividad del tipo Lección Magistral Tema 4 - Modelos de lenguaje Duración: 03:00 PL: Actividad del tipo Prácticas de Laboratorio			
10	Tema 4 - Modelos de lenguaje Duración: 03:00 LM: Actividad del tipo Lección Magistral Tema 4 - Modelos de lenguaje Duración: 03:00 PL: Actividad del tipo Prácticas de Laboratorio			

11				Realización de práctica TG: Técnica del tipo Trabajo en Grupo Evaluación continua Presencial Duración: 00:00
12				
13				
14				
15				
16				
17				Realización de práctica TI: Técnica del tipo Trabajo Individual Evaluación sólo prueba final No presencial Duración: 00:00

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
11	Realización de práctica	TG: Técnica del tipo Trabajo en Grupo	Presencial	00:00	100%	5 / 10	CB9 CB10 CT4 CB7 CT5 CE03 CB6 CB8 CT1 CT2 CT3 K01 K02 S03 S04 S05 CG1 CG2 CG3 CG4 CG5

7.1.2. Prueba evaluación global

Sem	Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
17	Realización de práctica	TI: Técnica del tipo Trabajo Individual	No Presencial	00:00	100%	5 / 10	CT4 CB7 CB9 CB10 CT5 CE03 CB6 CB8 CT1 CT2 CT3 K01 K02 S03

7.2. Criterios de evaluación

La evaluación, tanto si se sigue la opción de evaluación progresiva o de evaluación global, consistirá en la realización de una práctica al final de la asignatura que incluirá todo el temario. Su peso será del 100% y tendrá una nota mínima de 5.

Las prácticas en la evaluación progresiva se realizarán en parejas. En caso de suspender por evaluación progresiva, la práctica se realizará de forma individual en la evaluación global y extraordinaria.

Durante todas las sesiones se combinará la explicación de conocimientos teóricos con la realización de ejercicios prácticos obligatorios no evaluables.

8. Recursos didácticos

8.1. Recursos didácticos de la asignatura

Nombre	Tipo	Observaciones
Libro NLP Transformer	Bibliografía	Tunstall, Lewis, Leandro Von Werra, and Thomas Wolf. Natural language processing with transformers. " O'Reilly Media, Inc.", 2022.
Libro PyTorch	Bibliografía	Rao, Delip, and Brian McMahan. Natural language processing with PyTorch: build intelligent language applications using deep learning. " O'Reilly Media, Inc.", 2019.
Libro NLP con Python	Bibliografía	Hapke, Hannes, Cole Howard, and Hobson Lane. Natural Language Processing in Action: Understanding, analyzing, and generating text with Python. Simon and Schuster, 2019.

Libro NLP Systems	Bibliografía	Practical Natural Language Processing: A Comprehensive Guide to Building Real-World NLP Systems
-------------------	--------------	---