

Escuela de Doctorado y Estudios de Posgrado

Máster Universitario en Ciberseguridad e Inteligencia de Datos

GUÍA DOCENTE DE LA ASIGNATURA (ESCENARIO 0):

**Preprocesado de Datos
(2021 - 2022)**

1. Datos descriptivos de la asignatura

Asignatura: Preprocesado de Datos	Código: 835870904
<ul style="list-style-type: none"> - Centro: Escuela de Doctorado y Estudios de Postgrado - Lugar de impartición: - - Titulación: Máster Universitario en Ciberseguridad e Inteligencia de Datos - Plan de Estudios: 2018 (Publicado en 2018-09-19) - Rama de conocimiento: Ingeniería y Arquitectura - Itinerario / Intensificación: - Departamento/s: <ul style="list-style-type: none"> Matemáticas, Estadística e Investigación Operativa - Área/s de conocimiento: <ul style="list-style-type: none"> Estadística e Investigación Operativa - Curso: 1 - Carácter: - Duración: Primer cuatrimestre - Créditos ECTS: 3,0 - Modalidad de impartición: Semipresencial - Horario: Enlace al horario - Dirección web de la asignatura: http://www.campusvirtual.ull.es - Idioma: Castellano e Inglés (0,15 ECTS en Inglés) 	

2. Requisitos para cursar la asignatura

3. Profesorado que imparte la asignatura

Profesor/a Coordinador/a: CARLOS PEREZ GONZALEZ
- Grupo:
General <ul style="list-style-type: none"> - Nombre: CARLOS - Apellido: PEREZ GONZALEZ - Departamento: Matemáticas, Estadística e Investigación Operativa - Área de conocimiento: Estadística e Investigación Operativa
Contacto <ul style="list-style-type: none"> - Teléfono 1: 922318094 - Teléfono 2: - Correo electrónico: cpgonzal@ull.es - Correo alternativo: - Web: http://www.campusvirtual.ull.es
Tutorías primer cuatrimestre:

Desde	Hasta	Día	Hora inicial	Hora final	Localización	Despacho
Todo el cuatrimestre		Martes	16:00	19:00	Torre Profesor Agustín Arévalo - CE.1B	Despacho del profesor en la segunda planta de la torre.
Todo el cuatrimestre		Jueves	16:00	19:00	Torre Profesor Agustín Arévalo - CE.1B	Despacho del profesor en la segunda planta de la torre.

Observaciones: Las tutorías podrán variar dependiendo de la coincidencia con otras actividades académicas. El horario de tutorías se mantendrá actualizado en el aula virtual de la asignatura y los cambios en el mismo serán notificados a través de dicho medio.

Tutorías segundo cuatrimestre:

Desde	Hasta	Día	Hora inicial	Hora final	Localización	Despacho
Todo el cuatrimestre		Martes	16:00	19:00	Torre Profesor Agustín Arévalo - CE.1B	Despacho del profesor en la segunda planta de la torre.
Todo el cuatrimestre		Jueves	16:00	19:00	Torre Profesor Agustín Arévalo - CE.1B	Despacho del profesor en la segunda planta de la torre.

Observaciones: Las tutorías podrán variar dependiendo de la coincidencia con otras actividades académicas. El horario de tutorías se mantendrá actualizado en el aula virtual de la asignatura y los cambios en el mismo serán notificados a través de dicho medio.

Profesor/a: ARTURO JAVIER FERNANDEZ RODRIGUEZ

- Grupo:

General

- Nombre: **ARTURO JAVIER**
- Apellido: **FERNANDEZ RODRIGUEZ**
- Departamento: **Matemáticas, Estadística e Investigación Operativa**
- Área de conocimiento: **Estadística e Investigación Operativa**

Contacto

- Teléfono 1: **922 318179**
- Teléfono 2:
- Correo electrónico: **ajfernan@ull.es**
- Correo alternativo:
- Web: **http://www.campusvirtual.ull.es**

Tutorías primer cuatrimestre:

Desde	Hasta	Día	Hora inicial	Hora final	Localización	Despacho
Todo el cuatrimestre		Martes	11:00	14:00	Edificio de Física y Matemáticas - AN.2B	88
Todo el cuatrimestre		Miércoles	11:00	14:00	Edificio de Física y Matemáticas - AN.2B	88

Observaciones: El lugar y horario de tutorías pueden sufrir modificaciones puntuales que serán debidamente comunicadas en tiempo y forma.

Tutorías segundo cuatrimestre:

Desde	Hasta	Día	Hora inicial	Hora final	Localización	Despacho
Todo el cuatrimestre		Miércoles	13:00	15:00	Edificio de Física y Matemáticas - AN.2B	88
Todo el cuatrimestre		Jueves	10:30	14:30	Edificio de Física y Matemáticas - AN.2B	88

Observaciones: El lugar y horario de tutorías pueden sufrir modificaciones puntuales que serán debidamente comunicadas en tiempo y forma.

4. Contextualización de la asignatura en el plan de estudio

Bloque formativo al que pertenece la asignatura:
Perfil profesional:

5. Competencias

Generales

- CG1** - Ser capaces de aplicar los conocimientos adquiridos para resolver problemas en entornos nuevos o poco conocidos dentro de contextos relacionados con seguridad informática o inteligencia de datos
- CG3** - Mantener una actitud de permanente actualización, que les permita estudiar de manera autónoma mediante formación continua en su futuro desempeño profesional como expertos en seguridad informática e inteligencia de datos
- CG7** - Desarrollar las capacidades de trabajo en equipo y las habilidades de comunicación para mantener relaciones con otros profesionales y con organizaciones relevantes
- CG8** - Tener la capacidad analítica y de resolución para atender a los problemas reales de acuerdo con los valores éticos y sociales y con el máximo respeto a la legalidad vigente

Básicas

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

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

Específicas

CE12 - Conocer los métodos y procedimientos par la preparación de conjuntos de datos para aplicar técnicas de análisis de forma adecuada

6. Contenidos de la asignatura

Contenidos teóricos y prácticos de la asignatura

1. Preparación de datos (limpieza, transformación, normalización, imputación).
2. Reducción de datos (selección de características, selección de instancias, discretización, análisis de componentes principales ...)
3. Librerías y lenguajes de procesamiento de datos
 - Lenguaje R + Rstudio
 - Lenguaje Python+Jupyter

Actividades a desarrollar en otro idioma

En esta asignatura se impartirán 1,5 horas de clases en inglés.

En general, se trabajará preferentemente con bibliografía en inglés y el alumnado deberá ser capaz de extraer la información necesaria para seguir la asignatura a partir de dicha documentación, junto con los apuntes del profesorado. Asimismo, el software utilizado en prácticas y los manuales de uso y ayuda estarán en dicho idioma.

7. Metodología y volumen de trabajo del estudiante

Descripción

La metodología docente de las clases teóricas consistirá en sesiones en las que el profesorado explicará los conceptos fundamentales de cada tema que deben ser asimilados por el alumnado, bien presencialmente, o no presencialmente mediante retransmisión online, en directo usando videoconferencia o en diferido a través de grabaciones

colgadas en el entorno virtual.

La metodología docente de las clases prácticas consistirá en sesiones supervisadas en grupos reducidos en el laboratorio en las que se realizarán diversas prácticas informáticas de dificultad creciente aplicando los conceptos expuestos en las clases de teoría. Además, el alumnado aprenderá a usar diversas herramientas, en entornos reales o de simulación, así como metodologías relacionadas con el contexto de la materia.

La metodología docente de los informes consistirá en el desarrollo por parte del alumnado de su capacidad para la aplicación de los conocimientos adquiridos y la resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios y multidisciplinares, siendo capaces de integrar estos conocimientos.

Se plantea una metodología docente para los seminarios que consistirá en sesiones donde se llevará a cabo una explicación más detallada de determinados aspectos concretos de algunos temas teóricos o prácticos especialmente relevantes.

Se ofrecerán seminarios donde profesionales de esta materia harán charlas debates con el alumnado de los temas relacionados con el mundo profesional.

Las tutorías individuales ayudarán a reforzar los diferentes aspectos de la materia y ayudarán al alumnado en la comprensión de la teoría y la realización de las prácticas.

Actividades formativas en créditos ECTS, su metodología de enseñanza-aprendizaje y su relación con las competencias que debe adquirir el estudiante

Actividades formativas	Horas presenciales	Horas de trabajo autónomo	Total horas	Relación con competencias
Clases teóricas	5,00	5,00	10,0	[CB10], [CB8], [CG1]
Clases prácticas (aula / sala de demostraciones / prácticas laboratorio)	8,00	8,00	16,0	[CE12], [CB10], [CB8], [CG8], [CG1]
Realización de seminarios u otras actividades complementarias	1,00	1,00	2,0	[CB10], [CG3]
Estudio/preparación de clases teóricas	0,00	10,00	10,0	[CB10], [CB7], [CG3]
Estudio/preparación de clases prácticas	0,00	6,00	6,0	[CE12], [CB10], [CB7], [CG3]
Realización de exámenes	1,00	0,00	1,0	[CE12], [CB7], [CG1]
Asistencia a tutorías	0,00	6,00	6,0	[CB10], [CB7]
Informes, trabajos y proyectos	0,00	24,00	24,0	[CE12], [CB8], [CB7], [CG8], [CG7], [CG1]
Total horas	15,00	60,00	75,00	
		Total ECTS	3,00	

8. Bibliografía / Recursos

Bibliografía Básica

Salvador García, Julián Luengo, Francisco Herrera - Data Preprocessing in Data Mining [BBTK]

Mark Lutz, David Ascher - Learning Python [BBTK]

Edwin de Jonge, Mark van der Loo - An introduction to data cleaning with R [Rproject]

Hadley Wickham and Garrett Grolemund - R for data science : Import, tidy, transform, visualize, and model data [BBTK]

Bibliografía Complementaria

Otros Recursos

9. Sistema de evaluación y calificación

Descripción

La Evaluación de la asignatura se rige por el Reglamento de Evaluación y Calificación de la Universidad de La Laguna (BOC de 19 de enero de 2016), o el que la Universidad tenga vigente, además de por lo establecido en la Memoria de Verificación inicial o posteriores modificaciones

Modalidad de evaluación continua:

La evaluación continua de la asignatura se aplicará en las convocatorias ordinarias del curso y se llevará a cabo mediante la realización de un conjunto de actividades individuales y una prueba final. Las actividades individuales que los alumnos tendrán que llevar a cabo consistirán en la realización de cuestionarios, ejercicios y tareas entregables así como informes de prácticas. Los ejercicios serán calificados de 0-10 y

se aplicarán las siguientes ponderaciones en el cálculo de la nota por evaluación continua:

A: Prueba final (presentación de un proyecto sobre un supuesto práctico): 40% de la calificación final.

B: Cuestionarios, tareas y ejercicios: 20% de la calificación final.

C: Informes de prácticas: 40% de la calificación final.

La calificación final de la asignatura se calculará a partir de los pesos anteriores. Se deberá obtener una nota no inferior a 5 para superar la asignatura. La calificación obtenida en las actividades de los apartados B-C será válida para todas las convocatorias del curso académico, salvo que el(la) alumno(a) renuncie expresamente a ellas presentando la solicitud correspondiente (disponible en el aula virtual de la asignatura) al profesor responsable, con un mínimo de 6 días antes de la celebración de cualquiera de las convocatorias oficiales. La renuncia tendrá carácter definitivo para el resto de convocatorias del curso y el alumno optaría por la modalidad de evaluación alternativa.

Modalidad de evaluación alternativa:

Si un alumno renuncia o no pudiese desarrollar normalmente las actividades previstas en la evaluación continua, entonces deberá realizar las siguientes pruebas:

D: Prueba final escrita (examen+cuestionario): 60% de la calificación final.

E: Examen de prácticas basado en la resolución de un supuesto práctico: 40% de la calificación final.

La realización de estas pruebas tendrá lugar en las fechas oficiales de las convocatorias del curso y la calificación se calculará a partir de estos pesos, debiendo obtener una nota no inferior a 5 para superar la asignatura. La realización de estas pruebas será también la que se lleve a cabo en las convocatorias extraordinarias.

Estrategia Evaluativa

Tipo de prueba	Competencias	Criterios	Ponderación
Pruebas de respuesta corta	[CB10], [CB8], [CB7], [CG1]	- Nivel de conocimientos adquiridos y nivel de comprensión alcanzado en la aplicación de los contenidos explicados. - Adecuación a lo solicitado.	20,00 %
Trabajos y proyectos	[CE12], [CB10], [CB8], [CB7], [CG7], [CG3], [CG1]	- Calidad del trabajo desarrollado y concreción en la redacción - Interés demostrado - Adecuación a lo solicitado.	20,00 %
Informes memorias de prácticas	[CB10], [CB8], [CB7], [CG8], [CG3], [CG1]	- Nivel de comprensión alcanzado en la aplicación de los contenidos explicados así como la destreza técnica desarrollada durante las prácticas. - Adecuación a lo solicitado.	60,00 %

10. Resultados de Aprendizaje

El alumnado debe ser capaz de identificar qué técnica de preparación de datos o reducción de la dimensionalidad aplicar para adaptar los datos a los algoritmos de análisis, aumentar la eficiencia y precisión de los mismos o mejorar la comprensión del modelo obtenido.

11. Cronograma / calendario de la asignatura

Descripción

Debido al carácter semipresencial del máster, está previsto que las clases presenciales se desarrollen de la forma siguiente: El alumnado tendrá 3 horas diarias las semanas 1 a 5 del primer cuatrimestre.

Todas las asignaturas se desarrollarán en bimestres, y concretamente esta asignatura se impartirá en el bimestre 1. El siguiente cronograma recoge la distribución de actividades a modo orientativo. Se utilizará el aula virtual para informar del calendario de la asignatura y la distribución de las diferentes actividades presenciales programadas en la misma.

Primer cuatrimestre					
Semana	Temas	Actividades de enseñanza aprendizaje	Horas de trabajo presencial	Horas de trabajo autónomo	Total
Semana 1:	Tema 1	Clases teóricas y prácticas	2.00	0.00	2.00
Semana 2:	Tema 2	Clases teóricas y prácticas	3.00	7.00	10.00
Semana 3:	Tema 3	Clases teóricas y prácticas	3.00	7.00	10.00
Semana 4:	Tema 4	Clases teóricas y prácticas	3.00	7.00	10.00
Semana 5:	Tema 5	Clases teóricas y prácticas	3.00	7.00	10.00
Semana 6:	Temas 1 hasta 3	Actividades semipresenciales en el aula virtual. Realización de cuestionarios on-line. Videotutoriales y foro para la resolución de dudas.	0.00	10.00	10.00
Semana 7:	Tema 4	Actividades semipresenciales en el aula virtual. Realización de cuestionarios on-line. Videotutoriales y foro para la resolución de dudas.	0.00	10.00	10.00
Semana 8:	Tema 5	Actividades semipresenciales en el aula virtual. Realización de cuestionarios on-line. Videotutoriales y foro para la resolución de dudas.	0.00	10.00	10.00
Semana 16 a 18:	Evaluación	Evaluación del alumnado	1.00	2.00	3.00
Total			15.00	60.00	75.00