

Escuela de Doctorado y Estudios de Posgrado

Máster Universitario en Ciberseguridad e Inteligencia de Datos

GUÍA DOCENTE DE LA ASIGNATURA :

Extracción de Conocimiento en Bases de Datos (2019 - 2020)

1. Datos descriptivos de la asignatura

Asignatura: Extracción de Conocimiento en Bases de Datos	Código: 835870903
- 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: Ingeniería Informática y de Sistemas - Área/s de conocimiento: Ciencia de la Computación e Inteligencia Artificial - 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: JOSE MARCOS MORENO VEGA
- Grupo:
General - Nombre: JOSE MARCOS - Apellido: MORENO VEGA - Departamento: Ingeniería Informática y de Sistemas - Área de conocimiento: Ciencia de la Computación e Inteligencia Artificial
Contacto - Teléfono 1: 922318175 - Teléfono 2: - Correo electrónico: jmmoreno@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		Lunes	08:00	12:00	Edificio de Física y Matemáticas - AN.2B	84
Todo el cuatrimestre		Viernes	10:00	12:00	Edificio de Física y Matemáticas - AN.2B	84

Observaciones:

Tutorías segundo cuatrimestre:

Desde	Hasta	Día	Hora inicial	Hora final	Localización	Despacho
Todo el cuatrimestre		Martes	10:00	14:00	Edificio de Física y Matemáticas - AN.2B	84
Todo el cuatrimestre		Viernes	10:00	12:00	Edificio de Física y Matemáticas - AN.2B	84

Observaciones:

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

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

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

CE7 - Conocer las diferentes técnicas de análisis y explotación de datos y valorar su correcta aplicación en la extracción de información relevante

CE11 - Ser capaz de comunicar efectivamente las conclusiones alcanzadas tras analizar un conjunto de datos

6. Contenidos de la asignatura

Contenidos teóricos y prácticos de la asignatura

- Módulo 1.- Fundamentos del proceso de extracción de conocimiento desde fuentes de datos
 - Tema 1.- El proceso de extracción de conocimiento
- Módulo 2.- Aprendizaje supervisado. Modelos básicos de clasificación
 - Tema 2.- Árboles de decisión
 - Tema 3.- Clasificadores bayesianos
 - Tema 4.- Reglas de clasificación
 - Tema 5.- Evaluación de clasificadores
- Módulo 3.- Aprendizaje no supervisado. Modelos básicos de agrupamiento
 - Tema 6.- Agrupamiento basado en prototipos
 - Tema 7.- Agrupamiento basado en densidad
 - Tema 8.- Agrupamiento jerárquico
- Módulo 4.- Reglas de asociación
 - Tema 9.- Patrones frecuentes
 - Tema 10.- Medidas de evaluación
- Módulo 5.- Detección de anomalías
 - Tema 11.- Algoritmos basados en distancia y densidad

Actividades a desarrollar en otro idioma

En esta asignatura se impartirá 1,5 horas de clases en inglés. El software utilizado en las sesiones prácticas de laboratorio está documentado en inglés. También lo están los manuales y tutoriales que usarán los alumnos para el desarrollo de las prácticas asociadas a esta asignatura. Además, gran parte del material bibliográfico y de trabajo de la asignatura está escrito en este idioma. Se refuerza con ello la comprensión lectora de este idioma por parte de los alumnos.

Por otro lado, la memoria del proyecto (que se describe en el apartado Metodología y volumen del trabajo del estudiante) incluirá un breve resumen en inglés y los alumnos expondrán oralmente las conclusiones de su trabajo en este idioma. La evaluación de estas actividades se incluye en la evaluación del proyecto, según lo descrito en el apartado Sistema de evaluación y calificación.

7. Metodología y volumen de trabajo del estudiante

Descripción

La metodología docente de las clases teóricas consistirán en sesiones en las que el profesor explicará los conceptos fundamentales de cada tema que deben ser asimilados por el alumnado, 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 de laboratorio presenciales consistirán 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, trabajos y proyectos consistirán 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 multidisciplinarios, 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.

Para fomentar la comprensión del proceso de extracción de conocimiento y de las técnicas asociadas, se usará la metodología de enseñanza aprendizaje basada en proyectos. Se persigue con ello favorecer el aprendizaje efectivo, potenciar el trabajo autónomo, reforzar la motivación e implicación y favorecer la actitud reflexiva y crítica de los alumnos.

Desde fuentes suministradas por el profesor, los alumnos que tendrán que recopilar, preparar y tratar datos con el propósito de extraer conocimiento útil de los mismos. Además, deberán interpretar los resultados obtenidos y proponer estrategias para difundir y usar el conocimiento extraído.

El trabajo realizado se recogerá en una memoria del proyecto que deberá ser presentada oralmente al finalizar el cuatrimestre. El resumen del proyecto y las conclusiones derivadas el mismo se redactarán en inglés.

La memoria escrita y la exposición se usarán como base para la evaluación continua. Otras actividades complementarias a desarrollar por el alumno son la participación en foros, la búsqueda de información reciente sobre el tratamiento de datos y sus implicaciones legales, éticas y morales y la realización de ejercicios. Algunas de estas actividades se realizarán virtualmente.

La asignatura participa en el Programa de Apoyo a la Docencia Presencial mediante Herramientas TIC. Algunas de las tareas que se realizarán en el aula virtual son la participación en foros y la realización de cuestionarios.

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	[CB6], [CE7]
Clases prácticas (aula / sala de demostraciones / prácticas laboratorio)	8,00	8,00	16,0	[CG1], [CB7], [CE7]
Realización de seminarios u otras actividades complementarias	1,00	1,00	2,0	[CB6], [CE7]
Estudio/preparación de clases teóricas	0,00	10,00	10,0	[CG3], [CG7], [CB10], [CE7]
Estudio/preparación de clases prácticas	0,00	6,00	6,0	[CG3], [CG7], [CB10], [CE7]
Realización de exámenes	1,00	0,00	1,0	[CE7], [CE11]
Asistencia a tutorías	0,00	6,00	6,0	[CE7]
Informes, trabajos y proyectos	0,00	24,00	24,0	[CG8], [CB7], [CB8], [CE11]
Total horas	15,00	60,00	75,00	

Total ECTS	3,00	
------------	------	--

8. Bibliografía / Recursos

Bibliografía Básica

José Hernández Orallo, M.José Ramírez Quintana, Cesar Ferri Ramírez.
Introducción a la Minería de Datos
. Editorial Pearson, 2004

Ethem Alpaydin,
Introduction to Machine Learning
, The MIT Press, 2004

Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, Vipin Kumar,
Introduction to Data Mining
, Addison-Wesley, 2006

Charu. C. Aggarwal,
Data Mining.
The Textbook
. Springer, 2015

Bibliografía Complementaria

Ian H. Witten, Eibe Frank, Mark A. Hall, Christopher J. Pal,
Data Mining
. Practical Machine Learning Tools and Techniques
, Morgan Kaufmann, 2017

Foster Provost, Tom Fawcett, Data Science for Business, O'Reilly Media, 2013

Michael R. Berthold, Christian Borgelt, Frank Höppner, Frank Klawonn,
Guide to Intelligent Data Analysis
, Springer, 2010

Yanchang Zhao,
R and Data Mining: Examples and Case Studies,
Academic Press, 2012

Michael J. Crawley, The R Book, Wiley, 2013

Otros Recursos

R project

kaggle

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.

Se emplea un sistema de evaluación continua que contempla la realización de las siguientes actividades:

- el proyecto descrito en el apartado Metodología y volumen de trabajo del alumno,
- ejercicios prácticos individuales o en grupo, y
- participación en foros.

La calificación del proyecto tiene un peso del 80% de la nota (incluidas las actividades desarrolladas en inglés que se describen en el apartado Metodología y volumen de trabajo), correspondiendo el 20% restante a los ejercicios prácticos y a la participación en los foros.

Todas las actividades de la evaluación continua tienen carácter obligatorio.

En el apartado Cronograma/calendario de la asignatura se recogen las fechas estimadas de presentación de las diferentes actividades. Los resultados de las mismas serán comunicados a los alumnos aproximadamente 15 días después de su presentación.

Los alumnos que no superen la asignatura por el sistema de evaluación continua deberán presentarse, en las convocatorias y fechas establecidas para ello por la Universidad de La Laguna y la Escuela de doctorado y estudios de posgrado, a una prueba de evaluación alternativa. Esta consistirá en la realización de unos ejercicios prácticos y en la presentación y defensa, de manera individual, del proyecto descrito en el apartado Metodología y volumen de trabajo. La calificación de las actividades de carácter obligatorio obtenida en la evaluación continua se incorporará a la calificación final de la prueba alternativa

Estrategia Evaluativa

Tipo de prueba	Competencias	Criterios	Ponderación
----------------	--------------	-----------	-------------

Proyecto	• Memoria • Estructura, calidad y claridad de la redacción, fuentes consultadas, rigor en el análisis de los datos y coherencia de las conclusiones. • Defensa oral • Estructura de la exposición, lenguaje empleado y respuesta a las preguntas. • Análisis de datos • Grado de conocimiento adquirido en el manejo de las herramientas para el tratamiento inteligente de datos. • Rigor en el análisis de los datos y coherencia de las conclusiones.	80,00 %
Evaluación de teoría (pruebas individuales)	• -Frecuencia y pertinencia de las intervenciones en los foros. • Adecuación de las respuestas dadas a los ejercicios y claridad de redacción de las mismas.	20,00 %

10. Resultados de Aprendizaje

El alumnado debe ser capaz de extraer información relevante a partir de datos expresados en distintas formas. Para ello debe conocer, saber seleccionar y analizar resultados de aplicar técnicas básicas de minería de datos supervisadas y no supervisadas, incluyendo, clasificación y agrupamiento.

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 y 8 a 12 del primer cuatrimestre, y 3 o 4 horas diarias las semanas 1 a 5 del segundo cuatrimestre. Todas las asignaturas se desarrollarán en bimestres, y concretamente esta asignatura se impartirá en el bimestre 1. Sin embargo, las evaluaciones de todas las asignaturas deben ubicarse en las semanas 16 a 18 (en convocatorias oficiales).

El cronograma se presenta a título estimativo, de modo que el profesorado podrá modificar dicha planificación temporal si así lo demanda el desarrollo de la asignatura.

Primer cuatrimestre

Semana	Temas	Actividades de enseñanza aprendizaje	Horas de trabajo presencial	Horas de trabajo autónomo	Total
Semana 1:	Tema 1: El proceso de extracción de conocimiento en bases de datos	- Explicar el tema - Explicar las principales características del lenguaje R y sus paquetes - Tarea 1: Participar en el foro "Importancia, alcance, retos, oportunidades y límites del tratamiento de datos "	1.00	2.00	3.00
Semana 2:	Tema 2: Árboles de decisión	- Explicar el tema - Describir el proyecto que los alumnos deben realizar.	1.00	2.00	3.00
Semana 3:	Tema 2: Árboles de decisión	- Explicar el tema - Explicar los paquetes tree, randomForest de R - Tarea 2: Clasificación - Árboles de decisión (enunciado) - Desarrollo del proyecto	1.00	3.00	4.00
Semana 4:	Tema 3: Clasificadores bayesianos	- Explicar el tema - Explicar el paquete naivebayes de R - Tarea 2: Clasificación - Clasificador naive Bayes (enunciado) - Desarrollo del proyecto	1.00	4.00	5.00
Semana 5:	Tema 4: Reglas de clasificación	- Explicar el tema - Desarrollo del proyecto	1.00	5.00	6.00
Semana 6:	Tema 5: Evaluación de clasificadores	- Explicar el tema - Desarrollo del proyecto	1.00	4.00	5.00
Semana 7:	Tema 5: Evaluación de clasificadores	- Explicar el tema - Desarrollo del proyecto	0.00	5.00	5.00
Semana 8:	Tema 6: Agrupamiento basado en prototipos	- Explicar el tema - Explicar el paquete stats de R - Tarea 2 (Entrega para su evaluación) - Tarea 3: Agrupamiento – k means (enunciado) - Desarrollo del proyecto	2.00	4.00	6.00

Semana 9:	Tema 7: Agrupamiento basado en densidad	• Explicar el tema • Explicar el paquete dbscan de R • Tarea 3: Agrupamiento – Dbscan (enunciado) • Desarrollo del proyecto	1.00	6.00	7.00
Semana 10:	Tema 8: Agrupamiento jerárquico	• Explicar el tema • Explicar el paquete fastcluster de R • Tarea 3: Agrupamiento – Jerárquico (enunciado) • Desarrollo del proyecto	1.00	4.00	5.00
Semana 11:	Tema 9: Patrones frecuentes	• Explicar el tema • Tarea 3: Entrega para su evaluación • Tarea 4: Patrones frecuentes (enunciado) • Desarrollo del proyecto	1.00	4.00	5.00
Semana 12:	Tema 9: Patrones frecuentes	• Explicar el tema • Explicar el paquete arules de R • Desarrollo del proyecto	1.00	3.00	4.00
Semana 13:	Tema 10: Medidas de evaluación	• Explicar el tema • Desarrollo del proyecto	0.00	5.00	5.00
Semana 14:	Tema 11: Detección de anomalías	• Explicar el tema • Tarea 4. Entrega para su evaluación	1.00	4.00	5.00
Semana 15:	Seminario	• Celebración del seminario "Extracción de conocimiento"	1.00	1.00	2.00
Semana 16 a 18:	Evaluación	• Preparación y realización de las pruebas evaluativas (redacción de la memoria, presentación del proyecto, prueba única (en su caso))	1.00	4.00	5.00
Total			15.00	60.00	75.00