



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

105000169 - Minería De Datos

PLAN DE ESTUDIOS

10II - Grado En Ingeniería Informatica

CURSO ACADÉMICO Y SEMESTRE

2021/22 - Segundo 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..... | 5 |
| 7. Actividades y criterios de evaluación..... | 7 |
| 8. Recursos didácticos..... | 10 |
| 9. Otra información..... | 11 |

1. Datos descriptivos

1.1. Datos de la asignatura

| | |
|--|--|
| Nombre de la asignatura | 105000169 - Minería de Datos |
| No de créditos | 3 ECTS |
| Carácter | Optativa |
| Curso | Cuarto curso |
| Semestre | Octavo semestre |
| Período de impartición | Febrero-Junio |
| Idioma de impartición | Castellano |
| Titulación | 10II - Grado en Ingeniería Informática |
| Centro responsable de la titulación | 10 - Escuela Técnica Superior De Ingenieros Informaticos |
| Curso académico | 2021-22 |

2. Profesorado

2.1. Profesorado implicado en la docencia

| Nombre | Despacho | Correo electrónico | Horario de tutorías * |
|---|-----------------|---------------------------|--|
| Maria Concepcion Bielza Lozoya | 2210 | c.bielza@upm.es | M - 15:00 - 19:00 V - 15:00 - 17:00 |
| Pedro Maria Larrañaga Mugica (Coordinador/a) | 2208 | pedro.larranaga@upm.es | M - 15:00 - 19:00 V - 15:00 - 17:00 |

* 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

- Probabilidades Y Estadística I

3.2. Otros conocimientos previos recomendados para cursar la asignatura

El plan de estudios Grado en Ingeniería Informática no tiene definidos otros conocimientos previos para esta asignatura.

4. Competencias y resultados de aprendizaje

4.1. Competencias

CG-1/21 - Capacidad de resolución de problemas aplicando conocimientos de matemáticas, ciencias e ingeniería.

CG-13/CE55 - Capacidad de comunicarse de forma efectiva con los compañeros, usuarios (potenciales) y el público en general acerca de cuestiones reales y problemas relacionados con la especialización elegida.

CG-19 - Capacidad de usar las tecnologías de la información y la comunicación.

CG-2/CE45 - Capacidad para el aprendizaje autónomo y la actualización de conocimientos, y reconocimiento de su necesidad en el área de la informática.

CG-5 - Capacidad de gestión de la información.

CG-6 - Capacidad de abstracción, análisis y síntesis

Ce 12/16 - Conocer los campos de aplicación de la informática, y tener una apreciación de la necesidad de poseer unos conocimientos técnicos profundos en ciertas áreas de aplicación; apreciación del grado de esta necesidad en, por lo menos, una situación.

Ce 14/15 - Conocer el software, el hardware y las aplicaciones existentes en el mercado, así como el uso de sus elementos, y capacidad para familiarizarse con nuevas aplicaciones informáticas.

Ce 17 - Conocer los temas informáticos avanzados de modo que permita a los alumnos vislumbrar y entender las fronteras de la disciplina, por medio de la inclusión de experiencias de aprendizaje que dirigen a los alumnos desde los temas elementales a los temas avanzados o los temas de los que se nutren los novísimos desarrollos.

Ce 44 - Conocimiento de tecnologías punteras relevantes y su aplicación.

4.2. Resultados del aprendizaje

RA277 - Dado un problema real elegir la tecnología informática existente en el mercado mas apropiada para su solución y diseñar su desarrollo e integración, analizando la viabilidad de su solución, lo que se puede y no se puede conseguir a través del estado actual de desarrollo de la tecnología usada, y lo que se espera que avance en el futuro.

RA278 - Desarrollar la solución matemática y algorítmica mas apropiada a un problema informático que requiera un tratamiento especialmente complejo, analizando y exponiendo su viabilidad.

RA280 - Obtención de las competencias lingüísticas comunicativas (comprensión, expresión, etc.) habladas y escritas en entornos académicos/profesionales nacionales/internacionales.

RA281 - Obtención de las técnicas necesarias para la realización de un informe o memoria sobre un trabajo realizado en un entorno socio-lingüístico nacional/internacional.

5. Descripción de la asignatura y temario

5.1. Descripción de la asignatura

La Minería de Datos trata de construir sistemas informáticos cuando no existe experiencia humana o cuando no es fácilmente explicable. Así, los datos se transforman en conocimiento. Entre las muchas aplicaciones exitosas pueden citarse la neurociencia computacional, analítica deportiva, industria 4.0, el reconocimiento del habla o de texto manuscrito, navegación autónoma de robots, recuperación de información documental, filtrado cooperativo, sistemas de diagnóstico, análisis de microarrays de ADN, etc.

Este curso expone varios métodos de Minería de Datos desde el punto de vista práctico. El objetivo es enfrentarse a un conjunto de datos sobre los que construir modelos clasificatorios con la ayuda del software WEKA.

5.2. Temario de la asignatura

1. Introducción a la minería de datos
 - 1.1. Reconocimiento de patrones
 - 1.2. Preproceso de datos
2. Clasificación supervisada
 - 2.1. Métodos de evaluación
 - 2.2. Vecinos más cercanos
 - 2.3. Clasificadores Bayesianos y teoría de la información
 - 2.4. Regresión logística
 - 2.5. Árboles de clasificación
 - 2.6. Inducción de reglas
 - 2.7. Selección de variables
 - 2.8. Metaclasificadores
3. Clasificación no supervisada
 - 3.1. Métodos particionales
 - 3.2. Clasificación ascendente jerárquica
 - 3.3. Clustering probabilista

6. Cronograma

6.1. Cronograma de la asignatura *

| Sem | Actividad presencial en aula | Actividad presencial en laboratorio | Tele-enseñanza | Actividades de evaluación |
|-----|--|-------------------------------------|--|--|
| 1 | Clases teóricas del tema 1 Duración: 02:00 LM: Actividad del tipo Lección Magistral | | Clases teóricas del tema 1 Duración: 02:00 LM: Actividad del tipo Lección Magistral | |
| 2 | Clases teóricas del tema 2.1 Duración: 00:00 LM: Actividad del tipo Lección Magistral | | Clases teóricas del tema 2.1 Duración: 00:00 LM: Actividad del tipo Lección Magistral | |
| 3 | Clases teóricas del tema 2.2 Duración: 02:00 LM: Actividad del tipo Lección Magistral | | Clases teóricas del tema 2.2 Duración: 02:00 LM: Actividad del tipo Lección Magistral | |
| 4 | Clases teóricas del tema 2.3 Duración: 02:00 LM: Actividad del tipo Lección Magistral | | Clases teóricas del tema 2.3 Duración: 02:00 LM: Actividad del tipo Lección Magistral | |
| 5 | Clases teóricas del tema 2.4 Duración: 02:00 LM: Actividad del tipo Lección Magistral | | Clases teóricas del tema 2.4 Duración: 02:00 LM: Actividad del tipo Lección Magistral | |
| 6 | Clases teóricas del tema 2.5 Duración: 02:00 LM: Actividad del tipo Lección Magistral | | Clases teóricas del tema 2.5 Duración: 02:00 LM: Actividad del tipo Lección Magistral | |
| 7 | Clases teóricas del tema 2.6 Duración: 02:00 LM: Actividad del tipo Lección Magistral | | Clases teóricas del tema 2.6 Duración: 02:00 LM: Actividad del tipo Lección Magistral | |
| 8 | Clases teóricas del tema 2.7 Duración: 02:00 LM: Actividad del tipo Lección Magistral | | Clases teóricas del tema 2.7 Duración: 02:00 LM: Actividad del tipo Lección Magistral | |
| 9 | Clases teóricas del tema 2.8 Duración: 02:00 LM: Actividad del tipo Lección Magistral | | Clases teóricas del tema 2.8 Duración: 02:00 LM: Actividad del tipo Lección Magistral | |
| 10 | Trabajo 1 y exposición oral (tema 2) Duración: 02:00 OT: Otras actividades formativas | | Trabajo 1 y exposición oral (tema 2) Duración: 02:00 OT: Otras actividades formativas | |
| 11 | Trabajo 1 y exposición oral (tema 2) Duración: 02:00 OT: Otras actividades formativas | | Trabajo 1 y exposición oral (tema 2) Duración: 02:00 OT: Otras actividades formativas | Entrega y exposición oral del trabajo 1 sobre clasificación supervisada PI: Técnica del tipo Presentación Individual Evaluación continua Presencial Duración: 02:00 |
| 12 | Clases teóricas del tema 3 Duración: 02:00 LM: Actividad del tipo Lección Magistral | | Clases teóricas del tema 3 Duración: 02:00 LM: Actividad del tipo Lección Magistral | |
| 13 | Clases teóricas del tema 3 Duración: 02:00 LM: Actividad del tipo Lección Magistral | | Clases teóricas del tema 3 Duración: 02:00 LM: Actividad del tipo Lección Magistral | Test EX: Técnica del tipo Examen Escrito Evaluación continua Presencial Duración: 00:30 |

| | | | | |
|----|---|--|---|--|
| 14 | Trabajo 2 y exposición oral (tema 3) Duración: 02:00 OT: Otras actividades formativas | | Trabajo 2 y exposición oral (tema 3) Duración: 02:00 OT: Otras actividades formativas | |
| 15 | Trabajo 2 y exposición oral (tema 3) Duración: 02:00 OT: Otras actividades formativas | | Trabajo 2 y exposición oral (tema 3) Duración: 02:00 OT: Otras actividades formativas | Entrega y exposición oral del trabajo 2 sobre clasificación no supervisada PI: Técnica del tipo Presentación Individual Evaluación continua Presencial Duración: 02:00 |
| 16 | | | | |
| 17 | | | | Entrega y exposición oral de los trabajos 1 y 2 sobre clasificación supervisada y no supervisada, así como examen tipo test PI: Técnica del tipo Presentación Individual Evaluación sólo prueba final Presencial Duración: 02:00 |

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.

* El cronograma sigue una planificación teórica de la asignatura y puede sufrir modificaciones durante el curso derivadas de la situación creada por la COVID-19.

7. Actividades y criterios de evaluación

7.1. Actividades de evaluación de la asignatura

7.1.1. Evaluación continua

| Sem. | Descripción | Modalidad | Tipo | Duración | Peso en la nota | Nota mínima | Competencias evaluadas |
|------|--|--|------------|----------|-----------------|-------------|---|
| 11 | Entrega y exposición oral del trabajo 1 sobre clasificación supervisada | PI: Técnica del tipo Presentación Individual | Presencial | 02:00 | 33.36% | 4 / 10 | CG-2/CE45 CG-5 CG-19 CG-13/CE55 CG-1/21 Ce 12/16 Ce 14/15 Ce 17 Ce 44 |
| 13 | Test | EX: Técnica del tipo Examen Escrito | Presencial | 00:30 | 50% | 4 / 10 | CG-6 Ce 14/15 CG-1/21 CG-2/CE45 |
| 15 | Entrega y exposición oral del trabajo 2 sobre clasificación no supervisada | PI: Técnica del tipo Presentación Individual | Presencial | 02:00 | 16.66% | 4 / 10 | CG-1/21 CG-2/CE45 CG-5 CG-19 CG-13/CE55 Ce 12/16 Ce 14/15 Ce 17 Ce 44 |

7.1.2. Evaluación sólo prueba final

| Sem | Descripción | Modalidad | Tipo | Duración | Peso en la nota | Nota mínima | Competencias evaluadas |
|-----|---|--|------------|----------|-----------------|-------------|---|
| 17 | Entrega y exposición oral de los trabajos 1 y 2 sobre clasificación supervisada y no supervisada, así como examen tipo test | PI: Técnica del tipo Presentación Individual | Presencial | 02:00 | 100% | 4 / 10 | CG-2/CE45 CG-5 CG-1/21 CG-6 CG-19 CG-13/CE55 Ce 12/16 Ce 14/15 Ce 17 Ce 44 |

7.1.3. Evaluación convocatoria extraordinaria

No se ha definido la evaluación extraordinaria.

7.2. Criterios de evaluación

La calificación vendrá dada a partir de la evaluación de:

1. El trabajo escrito sobre clasificación supervisada (tema 2) proporcionado por el alumno y su exposición oral en el aula. Se utilizará el software WEKA. Este trabajo tendrá un peso del 33.34%.
2. El trabajo escrito sobre clasificación no supervisada (tema 3) proporcionado por el alumno y su exposición oral en el aula. Se utilizará el software WEKA. Este trabajo tendrá un peso del 16.66%.
3. El examen tipo test sobre los contenidos de la asignatura. Tendrá un peso del 50%.

Estos pesos se utilizan para la nota final siempre que el alumno se presente a todo y obtenga una nota mínima de 4 sobre 10 en cada una de las 3 partes.

Si no se superan los mínimos previamente mencionados en alguna de las tres partes, deberá presentarse de nuevo a esa(s) parte(s) en Julio. Si, en cambio, todas superan los mínimos establecidos, pero la nota final no resulta un aprobado (mayor o igual que 5 sobre 10), el alumno tendrá que presentarse en Julio a cualquiera de las partes (una, dos o las tres), lo que él decida, teniendo en cuenta que para aprobar la nota final ha de ser mayor o igual a 5 sobre 10.

Para la convocatoria extraordinaria de Julio las evaluaciones serán análogas (dos exposiciones orales y sus respectivos trabajos escritos realizados de forma individual, junto con el examen de tipo test).

Si el estudiante decide realizar sólo una prueba final, se efectuará en la fecha de examen final fijada por jefatura de estudios y consistirá (como en la evaluación continua) en las dos exposiciones orales y sus respectivos trabajos escritos realizados de forma individual, y en el examen de tipo test, evaluándose con los mismos pesos que en la evaluación continua. La solicitud para hacer sólo una prueba final se tiene que realizar por email al coordinador de la asignatura antes de la primera prueba.

La situación sanitaria causada por la pandemia COVID-19 obliga a restringir el aforo de las aulas y por ello se ha decidido que la docencia de este semestre sea de presencialidad mixta. Se establecerán turnos de presencialidad dentro de los grupos, de forma que cada semana un turno asistirá a clase en el aula (columna "actividad en el aula" del cronograma), mientras el resto de los turnos se conectarán a la clase en remoto (columna "teleenseñanza").

Y cada semana será un turno diferente el que acuda al aula. Si mejoraran las condiciones sanitarias y se pudieran impartir clases presenciales con normalidad, todos los alumnos acudirán a las aulas a recibir las clases indicadas en la columna "actividad en el aula". Si, por el contrario, empeoraran las condiciones sanitarias, todos los alumnos pasarían a conectarse a las clases en remoto de la columna "tele-enseñanza". En esta situación las pruebas de evaluación continua presenciales previstas se realizarían de forma online, sin necesidad de modificar esta guía.

No podrá haber dos alumnos que utilicen los mismos datos y con el mismo objetivo.

Está terminantemente prohibido copiar o dejar copiar. Su detección implicará el suspenso inmediato y la repetición de los trabajos con datos diferentes.

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 |
|-----------------------|--------------|--|
| Libro 1 | Bibliografía | E. Alpaydin. Introduction to Machine Learning. MIT Press |
| Libro 3 | Bibliografía | J. Hernández-Orallo, M.J. Ramírez, C. Ferri. Introducción a la Minería de Datos. Pearson Educación. 2004. |
| Libro 4 | Bibliografía | L. Kuncheva. Combining Pattern Classifiers. Wiley. 2004 |
| Libro 7 | Bibliografía | I. Witten, E. Frank. Data Mining. Morgan Kaufmann. 2ª ed. 2005. |
| Libro 8 | Bibliografía | C. Bielza, P. Larrañaga (2019) Data-Driven Computational Neuroscience. Machine Learning and Statistical Models. Cambridge University Press |
| software | Recursos web | Software WEKA: http://www.cs.waikato.ac.nz/ml/weka |
| Repositorios de datos | Recursos web | - http://www.inside-r.org/howto/finding-data-internet - https://www.quora.com/Where-can-I-find-large-datasets-open-to-the-public |

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.

Se plantea docencia presencial por turnos, de forma que todas las semanas hay estudiantes en clase y también estudiantes en casa. Por ello en el cronograma aparecen duplicadas las actividades docentes en las dos columnas.

El alumno que tenga que realizar una presentación oral (cuando le toque) lo hará presencialmente en el aula. Para el test escrito todos los alumnos estarán presentes en el aula (o aulas).