GUIA DOCENTE DE LA ASIGNATURA (∞)

MINERÍA DE DATOS

Curso 2018-2019

(Fecha última actualización: 15/05/2018)
(Fecha de aprobación en Consejo de Departamento: 16/05/2018)
(Fecha de aprobación en Consejo de Departamento: de Ciencias de la Computación e I. A.: 23/05/2018)

MÓDULO	MATERIA	CURSO	SEMESTRE	CRÉDITOS	TIPO
Análisis Exploratorio y Minería de Datos	Análisis Exploratorio y Minería de Datos	4º	2º	6	Optativa

PROFESORES[1]	DIRECCIÓN COMPLETA DE CONTACTO PARA TUTORÍAS (Dirección postal, teléfono, correo electrónico, etc.)
• José Fernando Vera: Parte I "Técnicas Estadísticas para la Minería de Datos" • Mª Carmen Pegalajar: Parte II "Técnicas Informáticas para la Minería de Datos" • Joaquín Abellan: Parte II "Técnicas Informáticas para la Minería de Datos"	Departamento de Estadística e I. O. Facultad de Ciencias Universidad de Granada Campus Fuentenueva s/n 18071- Granada Teléfono: 958243157; Fax: 958243267 http://www.stei.es/estadistica/ jfvera@ugr.es Departamento de Ciencias de la Computación e I.A. E.T.S.I.I.T. - Universidad de Granada C/Daniel Saucedo Aranda s/n 18071-GRANADA Teléfono: 958244019; Fax: 948243317 http://decsai.ugr.es mcarmen@decsai.ugr.es jabellan@decsai.ugr.es
	HORARIO DE TUTORÍAS Y/O ENLACE A LA PÁGINA WEB DONDE PUEDAN CONSULTARSE LOS HORARIOS DE TUTORÍAS[1]
	L 10-12 horas; M 9-10, 12-14 horas; X 9-10 horas, profesor José Fernando Vera .

[1] Consulte posible actualización en Acceso Identificado > Aplicaciones > Ordenación Docente
(∞) Esta guía docente debe ser cumplimentada siguiendo la "Normativa de Evaluación y de Calificación de los estudiantes de la Universidad de Granada" (http://secretariageneral.ugr.es/pages/normativa/fichasugr/ncg7121/!)

Firmado por: RAMON GUTIERREZ SANCHEZ Secretario/a de Departamento
Sello de tiempo: 24/09/2019 17:21:59 Página: 1 / 8
2CSw1l0p104P9cHm7+IlgH5CKCJ3NmbA
La integridad de este documento se puede verificar en la dirección https://sede.ugr.es/verifirma/pfinicio.jsp introduciendo el código de verificación que aparece debajo del código de barras.

	Los horarios de los profesores María del Carmen Pegalajar y Joaquín Abellan pueden consultarse en la web: http://decsai.ugr.es/index.php?p=profesores
GRADO EN EL QUE SE IMPARTE	OTROS GRADOS A LOS QUE SE PODRÍA OFERTAR
Grado en Estadística	Matemáticas, Biología, Telecomunicaciones, Ingeniería Civil, Farmacia, Psicología, Medicina

PRERREQUISITOS Y/O RECOMENDACIONES (si procede)

- Aunque no es necesario que los alumnos tengan aprobadas asignaturas, materias o módulos previos como requisito indispensable para cursar este módulo, se recomienda la realización previa de los módulos Estadística y Estadística Computacional, así como la asignatura Análisis Exploratorio de Datos.

BREVE DESCRIPCIÓN DE CONTENIDOS (SEGÚN MEMORIA DE VERIFICACIÓN DEL GRADO)

- Técnicas de Minería de Datos para problemas de agrupamiento.
- Técnicas de Minería de Datos para problemas de clasificación.
- Técnicas de Minería de Datos para problemas de asociación.
- Herramientas software para Minería de Datos.

COMPETENCIAS GENERALES Y ESPECÍFICAS

Competencias generales:

G01. Poseer los conocimientos básicos de los distintos módulos que, partiendo de la base de la educación secundaria general, y apoyándose en libros de texto avanzados, se desarrollan en la propuesta de título de Grado en Estadística que se presenta.

G02. Saber aplicar los conocimientos básicos de cada módulo a su trabajo o vocación de una forma profesional y poseer las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y la resolución de problemas dentro de la Estadística y ámbitos en que esta se aplica directamente.

G03. Saber reunir e interpretar datos relevantes para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética.

G04. Poder transmitir información, ideas, problemas y sus soluciones, de forma escrita u oral, a un público tanto especializado como no especializado.

G05. Haber desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía.

G06. Saber utilizar herramientas de búsqueda de recursos bibliográficos.

G08. Poseer habilidades y aptitudes que favorezcan el espíritu emprendedor en el ámbito de aplicación y

Firmado por: RAMON GUTIERREZ SANCHEZ Secretario/a de Departamento
Sello de tiempo: 24/09/2019 17:21:59 Página: 2 / 8
2CSw1l0p104P9cHm7+IlgH5CKCJ3NmbA
La integridad de este documento se puede verificar en la dirección https://sede.ugr.es/verifirma/pfinicio.jsp introduciendo el código de verificación que aparece debajo del código de barras.

desarrollo de su formación académica.

Competencias específicas:

E01. Conocer los fundamentos básicos del razonamiento estadístico, en el diseño de estudios, en la recogida de información, en el análisis de datos y en la extracción de conclusiones.

E02. Conocer, saber seleccionar y saber aplicar, técnicas de adquisición de datos para su tratamiento estadístico.

E03. Conocer los fundamentos teóricos y saber aplicar modelos y técnicas estadísticas en estudios y problemas reales en diversos ámbitos científicos y sociales.

E04. Saber seleccionar los modelos o técnicas estadísticas para su aplicación en estudios y problemas reales en diversos ámbitos científicos y sociales, así como conocer herramientas de validación de los mismos.

E07. Conocer los conceptos y herramientas matemáticas necesarias para el estudio de los aspectos teóricos y prácticos de la Probabilidad, la Estadística y la Investigación Operativa.

E08. Conocer y saber utilizar aplicaciones informáticas de análisis estadístico, cálculo numérico y simbólico, bases de datos, visualización gráfica y optimización, que sean útiles para la aplicación y desarrollo de las técnicas estadísticas.

E10. Tomar conciencia de la necesidad de asumir las normas de ética profesional y las relativas a la protección de datos y del secreto estadístico, como premisas que deben guiar la actividad profesional como profesionales de la Estadística.

OBJETIVOS (EXPRESADOS COMO RESULTADOS ESPERABLES DE LA ENSEÑANZA)

- Comprender la necesidad de una exploración previa de los datos, su preparación y adecuación, así como conocer las técnicas de filtrado y transformación.
- Comprender el concepto de la minería de datos (MD), su relación con otras áreas de los campos de la Estadística y de la Informática, y su papel en los procesos de extracción de conocimiento y ayuda a la toma de decisiones.
- Conocer los problemas que pueden ser resueltos mediante MD, qué técnicas son aplicables en cada caso y cuáles son sus limitaciones.
- Conocer las técnicas más habituales en MD, correspondientes a agrupamientos, clasificación asociación y predicción.
- Ser capaz de utilizar adecuadamente las herramientas software para MD.

TEMARIO DETALLADO DE LA ASIGNATURA

TEMARIO TEÓRICO:

Tema 1: Perspectiva Estadística de la Minería de Datos.
- La minería de datos como técnica estadística de exploración y validación.
- Problemas y técnicas estadísticas en la minería de datos. Big Data.

Tema 2. Minería de datos mediante técnicas estadísticos de agrupamiento.
- Medidas de proximidad.
- Criterios de análisis cluster para matrices de disimilaridad.
- Agrupamiento mediante mezclas de distribuciones.

Firmado por: RAMON GUTIERREZ SANCHEZ Secretario/a de Departamento
Sello de tiempo: 24/09/2019 17:21:59 Página: 3 / 8
2CSw1l0p104P9cHm7+IlgH5CKCJ3NmbA
La integridad de este documento se puede verificar en la dirección https://sede.ugr.es/verifirma/pfinicio.jsp introduciendo el código de verificación que aparece debajo del código de barras.

-Aplicaciones de minería de datos mediante técnicas de agrupamiento con SPSS y R.

Tema 3. Minería de datos mediante técnicas estadísticas de clasificación.
-Clasificación mediante análisis discriminante.
-Clasificación mediante discriminación logística.
-Otros métodos de clasificación.
-Aplicaciones de minería de datos mediante técnicas de clasificación con SPSS y R.

Tema 4: Perspectiva informática de la Minería de Datos. Modelos de asociación.
- La Minería de Datos y su relación con la Informática
- Selección y transformación de datos.
- Concepto de patrón frecuente y regla de asociación o tendencia
- Métodos para la obtención de reglas de asociación
- Medidas de interés para la evaluación del modelo

Tema 5: Modelos de Agrupamiento o Segmentación
- Modelos de agrupamiento con un número fijo de grupos
- Agrupamiento jerárquico
- Evaluación de modelos

Tema 6: Modelos Predictivos
- Regresión
- Modelos de Clasificación:
  * Basados en instancias (lazy)
  * Bayesianos
  * Basados en árboles de decisión
- Evaluación del clasificador
- Clasificación con costes

TEMARIO PRÁCTICO:

Práctica 1: Aplicación de técnicas estadísticas de la Minería de Datos.
-Metodología del análisis estadístico computacional para la Minería de Datos con SPSS.
-Metodología del análisis estadístico computacional para la Minería de Datos con R.

Práctica 2: Aplicación de las técnicas estadísticas de Agrupamiento.
-Aplicación del análisis cluster. Algoritmo de K-medias.
-Aplicación de los modelos de clases latentes.

Práctica 3:Aplicación de las técnicas estadísticas de Clasificación.
-Aplicación del análisis discriminante.
-Aplicación de la regresión logística.

Práctica 4: Herramientas para la Minería de Datos. Modelos de asociación
- Técnicas y Herramientas para selección de datos
- Técnicas y Herramientas para transformación de datos.
- Knime-Weka
- Detección de patrones frecuentes y tendencias
- Filtrado de reglas y aplicación de medidas de interés

Práctica 5. Creación de modelos de agrupamiento y segmentación

Firmado por: RAMON GUTIERREZ SANCHEZ Secretario/a de Departamento
Sello de tiempo: 24/09/2019 17:21:59 Página: 4 / 8

2CSw1l0p104P9cHm7+IlgH5CKCJ3NmbA

La integridad de este documento se puede verificar en la dirección https://sede.ugr.es/verifirma/pfinicio.jsp introduciendo el código de verificación que aparece debajo del código de barras.

- Con valores de k (número de grupos) fijados de antemano
- Con conjuntos de datos artificiales creados por el alumno conteniendo datos esféricos, no esféricos y con/sin outliers
- Sin fijar un valor de k, sino obteniéndolo a través de un diagrama SSE

Práctica 6. Creación de modelos de predicción
- Con variables numéricas (regresión)
- Con variables dependientes nominales (clasificación)
- Ajuste del modelo, usando métodos como validación cruzada

BIBLIOGRAFÍA

BIBLIOGRAFÍA FUNDAMENTAL:

1. Multivariate Analysis. Mardia, K.V., Kent, J.T., and Bibby, J.M. Academic Press. 1979
2. Handbook of Computational Statistics. Concepts and Methods. Gentle, J.E., Härdle, W., Mori, Y Editors. Springer. 2004.
3. Applied Data Mining: Statistical Methods for Business and Industry. P. Guidici. Wiley, 2003
4. Introduction to Data Mining. Pang-Ning Tan, Michael Steinbach, Vipin Kumar. Addison Wesley, 2005.
5. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition. Trevor Hastie
6. Trevor Hastie, Robert Tibshirani, Jerome Friedman (Author) . Springer; 2nd ed. 2009
7. Introducción a la Minería de Datos. José Hernández Orallo, M.José Ramírez Quintana, Cèsar Ferri Ramírez. Pearson, 2004.

BIBLIOGRAFÍA COMPLEMENTARIA:

1. Handbook of Statistical Analysis and Data Mining Applications. Robert Nisbet, John Elder IV, Gary Miner. Academic Press; 2009.
2. Data Mining: Practical Machine Learning Tools and Techniques. Ian H. Witten, Eibe Frank. Morgan Kaufmann, 2005
3. Data Mining, Second Edition: Concepts and Techniques. Jiawei Han, Micheline Kamber, Jian Pei. Morgan Kaufmann, 2006

ENLACES RECOMENDADOS

http://elvex.ugr.es/idbis/dm/ (Página web del grupo de investigación IDBIS del departamento de Ciencias de la Computación e Inteligencia Artificial, con material docente y de investigación sobre Minería de Datos)
http://www.kdnuggets.com/ (Principal página en la Web sobre todo tipo de recursos relacionados con la Minería de Datos)

Firmado por: RAMON GUTIERREZ SANCHEZ Secretario/a de Departamento
Sello de tiempo: 24/09/2019 17:21:59 Página: 5 / 8
2CSw1l0p104P9cHm7+IlgH5CKCJ3NmbA
La integridad de este documento se puede verificar en la dirección https://sede.ugr.es/verifirma/pfinicio.jsp introduciendo el código de verificación que aparece debajo del código de barras.

http://www.r-project.org/ (Página principal de la plataforma R)

METODOLOGÍA DOCENTE

- Clases de teoría: Sesiones para todo el grupo de alumnos en las que se explicarán, por parte del profesor, los contenidos teóricos fundamentales y su importancia en el contexto de la materia, y que servirán para fijar los conocimientos ligados a las competencias previstas.

- Clases de problemas y de prácticas en ordenador: Sesiones para todo el grupo de alumnos en las que el profesor realizará ejercicios y problemas sobre los contenidos teóricos de cada tema y guiará a los alumnos en la resolución de otros similares. También, dependiendo de las materias, y mediante la utilización de programas de ordenador adecuados, el profesor realizará prácticas en ordenador para temas específicos y guiará a los alumnos en el planteamiento y realización de nuevas prácticas. Así, tanto en las clases de problemas como en las clases de prácticas en ordenador, se aplicarán los contenidos adquiridos en las teóricas, de modo que los estudiantes vayan obteniendo las competencias previstas. Para las clases de prácticas en ordenador se harán subgrupos de tamaño acorde con las posibilidades de las aulas de informática.

- Seminarios y trabajos tutelados: Como complemento de las clases teóricas y prácticas, los profesores podrán proponer a los estudiantes la realización de trabajos individuales o en grupos reducidos, para lo que podrán contar con el apoyo del profesor en tutorías. En los seminarios, organizados para todo el grupo de alumnos, se realizarán debates sobre la materia, se discutirán aspectos específicos del temario, y los alumnos desarrollarán ejercicios y trabajos, y podrán compartir con sus compañeros y con el profesor las dudas que encuentren y obtener solución a las mismas.

- Actividades no presenciales individuales (Estudio y trabajo autónomo). Los estudiantes habrán de llevar a cabo una tarea personal de estudio y asimilación de la teoría y de preparación y resolución de trabajos y problemas propuestos, para alcanzar las competencias previstas.

- Actividades no presenciales grupales. Además, los estudiantes habrán de llevar a cabo la realización de trabajos en grupo, para lo que podrán contar con el apoyo del profesor en Tutorías académicas, de forma que los estudiantes puedan compartir con sus compañeros y con el profesor las dudas que encuentren, obtener solución a las mismas y comenzar a alcanzar por sí mismos las competencias de la materia.

- Tutorías académicas. Ofrecerán apoyo y asesoramiento, personalizado o en grupos formados por un pequeño número de alumnos, para abordar las tareas encomendadas en las actividades formativas indicadas previamente o específicas del trabajo personal. El profesor jugará un papel proactivo, orientando hacia un aprendizaje cooperativo, a lo largo de todo el curso.

Las anteriores actividades formativas se desarrollarán desde una metodología participativa y aplicada centrada en el trabajo del estudiante (presencial y no presencial / individual y grupal) según la siguiente distribución aproximada (atendiendo a las indicaciones generales de los módulos del grado):
- Un 40% de docencia presencial en el aula.
- Un 50% de estudio individualizado del alumno, búsqueda, consulta y tratamiento de información, resolución de problemas y casos prácticos, y realización de trabajos y exposiciones.
- Un 10% para tutorías colectivas y evaluación.

Para esta materia, la relación de cada actividad formativa con las competencias a adquirir es la siguiente:
Clases de teoría

Firmado por: RAMON GUTIERREZ SANCHEZ Secretario/a de Departamento
Sello de tiempo: 24/09/2019 17:21:59 Página: 6 / 8
2CSw1l0p104P9cHm7+IlgH5CKCJ3NmbA
La integridad de este documento se puede verificar en la dirección https://sede.ugr.es/verifirma/pfinicio.jsp introduciendo el código de verificación que aparece debajo del código de barras.

Competencias generales: G01, G03, G05.
Competencias específicas: E01, E02, E03, E04, E06, E08, E09.

Clases de problemas y prácticas en ordenador
Competencias generales: G01, G02.
Competencias específicas: E02, E04, E06, E08, E09.

Seminarios y exposición de trabajos
Competencias generales: G01, G02, G03, G04, G05, G06, G08.
Competencias específicas: E04, E06, E08, E09.

Tutorías
Competencias generales: G01, G03, G04, G05.
Competencias específicas: E01, E02, E03, E04, E06, E08, E09.

Trabajo personal del alumno
Competencias generales: G01, G02, G03, G04, G05, G06, G08.
Competencias específicas: E01, E02, E03, E04, E06, E08, E09.

EVALUACIÓN (INSTRUMENTOS DE EVALUACIÓN, CRITERIOS DE EVALUACIÓN Y PORCENTAJE SOBRE LA CALIFICACIÓN FINAL, ETC.)

Todo lo relativo a la evaluación se regirá por la normativa de planificación docente y organización de exámenes de la Universidad de Granada.

El sistema de calificación empleado será el establecido en el artículo 5 del Real Decreto 1125/2003, de 5 de septiembre, por el que se establece el sistema europeo de créditos y el sistema de calificaciones en las titulaciones universitarias de carácter oficial y validez en todo el territorio nacional.

Los criterios de evaluación se indicarán en los Programas y Guías Didácticas correspondientes a cada asignatura, garantizando así la transparencia y objetividad de los mismos.
De acuerdo con el Real Decreto 1125/2003, la valoración del nivel de adquisición de las competencias generales y específicas de cada materia se llevará a cabo de manera continua a lo largo de todo el periodo académico.

La evaluación se realizará a partir de la medición de las diversas actividades que realizan los alumnos. La superación de cualquiera de las pruebas no se logrará sin un conocimiento uniforme y equilibrado de toda la materia. Se tendrán en cuenta los siguientes procedimientos, aspectos y criterios, asignando a cada uno de ellos un porcentaje que se mantenga en el rango indicado y de tal manera que la suma de los tres constituya el total (100%) de la calificación:

• Pruebas específicas de conocimientos y resolución de ejercicios, orales y escritas, donde se valorarán tanto la asimilación como la expresión de los conocimientos adquiridos por el alumno, especialmente, su capacidad para la aplicación de los mismos a situaciones prácticas concretas y se realizará una observación sistemática del proceso de aprendizaje. Aportará el 45% de la calificación.

• Trabajos y seminarios. Abarca todos los trabajos y seminarios realizados por los estudiantes a lo largo del curso (ejercicios, prácticas en ordenador, resúmenes, cuadernos de trabajo, presentaciones, entrevistas, cuestionarios, etc.), tanto de carácter individual como en grupo. Se valorará además de los propios trabajos, la presentación y defensa de los mismos, y los debates suscitados en los seminarios, para obtener información sobre aspectos actitudinales, de integración y actuación social. Aportará el 50% de la calificación.

• Participación, actitud y esfuerzo personal de los alumnos en todas las actividades formativas programadas,

Firmado por: RAMON GUTIERREZ SANCHEZ Secretario/a de Departamento
Sello de tiempo: 24/09/2019 17:21:59 Página: 7 / 8
2CSw1l0p104P9cHm7+IlgH5CKCJ3NmbA
La integridad de este documento se puede verificar en la dirección https://sede.ugr.es/verifirma/pfinicio.jsp introduciendo el código de verificación que aparece debajo del código de barras.

así como una autoevaluación razonada. Aportará el 5% de la calificación.
DESCRIPCIÓN DE LAS PRUEBAS QUE FORMARÁN PARTE DE LA EVALUACIÓN ÚNICA FINAL ESTABLECIDA EN LA "NORMATIVA DE EVALUACIÓN Y DE CALIFICACIÓN DE LOS ESTUDIANTES DE LA UNIVERSIDAD DE GRANADA"
La evaluación única final establecida en la Normativa de evaluación y de calificación de los estudiantes de la Universidad de Granada consistirá en un examen escrito en el que se incluirán preguntas teóricas y prácticas sobre el temario que figura en esta guía docente.
INFORMACIÓN ADICIONAL
Cumplimentar con el texto correspondiente en cada caso

Firmado por: RAMON GUTIERREZ SANCHEZ Secretario/a de Departamento

Sello de tiempo: 24/09/2019 17:21:59 Página: 8 / 8

2CSw1l0p104P9cHm7+IlgH5CKCJ3NmbA

La integridad de este documento se puede verificar en la dirección https://sede.ugr.es/verifirma/pfinicio.jsp introduciendo el código de verificación que aparece debajo del código de barras.