

Guía docente de la asignatura

Minería de Datos (22311C2)

Fecha de aprobación:

Departamento de Ciencias de la Computación e
Inteligencia Artificial: 14/06/2022Departamento de Estadística e Investigación Operativa:
20/06/2022

| | | | | | | | |
|---------------|--|-----------------|--|-----------------|---|-------------|----------|
| Grado | Grado en Estadística | Rama | Ciencias | | | | |
| Módulo | Análisis Exploratorio y Minería de Datos | Materia | Análisis Exploratorio y Minería de Datos | | | | |
| Curso | 4 ^o | Semestre | 2 ^o | Créditos | 6 | Tipo | Optativa |

PRERREQUISITOS Y/O RECOMENDACIONES

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 ASOCIADAS A MATERIA/ASIGNATURA**COMPETENCIAS GENERALES**

- CG01 - CG01. 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.
- CG02 - CG02. 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.
- CG03 - CG03. Saber reunir e interpretar datos relevantes para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética.
- CG04 - CG04. Poder transmitir información, ideas, problemas y sus soluciones, de forma escrita u oral, a un público tanto especializado como no especializado.
- CG05 - CG05. Haber desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía.



- CG06 - CG06. Saber utilizar herramientas de búsqueda de recursos bibliográficos.
- CG08 - CG08. Poseer habilidades y aptitudes que favorezcan el espíritu emprendedor en el ámbito de aplicación y desarrollo de su formación académica.

COMPETENCIAS ESPECÍFICAS

- CE01 - CE01. 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.
- CE02 - CE02. Conocer, saber seleccionar y saber aplicar, técnicas de adquisición de datos para su tratamiento estadístico.
- CE03 - CE03. Conocer los fundamentos teóricos y saber aplicar modelos y técnicas estadísticas en estudios y problemas reales en diversos ámbitos científicos y sociales.
- CE04 - CE04. 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.
- CE06 - CE06. Comprender y utilizar básicamente el lenguaje matemático.
- CE08 - CE08. 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.
- CE09 - CE09. Conocer los conceptos básicos y habilidades propias de un ámbito científico o social en el que la Estadística o la Investigación operativa sean una herramienta fundamental.

RESULTADOS DE APRENDIZAJE (Objetivos)

- 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.

PROGRAMA DE CONTENIDOS TEÓRICOS Y PRÁCTICOS

TEÓRICO

Tema 1: Perspectiva Estadística de la Minería de Datos.

- La minería de datos como técnica estadística de exploración y validación.
- Técnicas estadísticas de preprocesamiento de datos. Detección de datos anómalos. Reducción de dimensiones. Imputación de datos faltantes.

Tema 2: Minería de datos mediante técnicas estadísticas de agrupamiento.



- Medidas de proximidad.
- Agrupamiento mediante el método K-medias
- Agrupamiento mediante mezclas de distribuciones.
- Aplicaciones de minería de datos mediante técnicas de agrupamiento con R.

Tema 3. Minería de datos mediante técnicas estadísticas de clasificación.

- Clasificación mediante regresión logística.
- Clasificación mediante análisis discriminante.
- Evaluación de modelos de clasificación.
- Aplicaciones de minería de datos mediante técnicas de clasificación con 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

PRÁCTICO

Práctica 1: Aplicación de técnicas estadísticas de la Minería de Datos.

- 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, Python, R
- Detección de patrones frecuentes y tendencias
- Filtrado de reglas y aplicación de medidas de interés

Práctica 5. Creación de modelos de agrupamiento y segmentación

- 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. An Introduction to Statistical Learning with Applications in R. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani. Springer, 2013.
6. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition. Trevor Hastie, Robert Tibshirani, Jerome Friedman. 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)
- <http://www.r-project.org/> (Página principal de la plataforma R)

METODOLOGÍA DOCENTE

- MD01 – MD1. Lección magistral/expositiva
- MD02 – MD2. Sesiones de discusión y debate
- MD03 – MD3. Resolución de problemas y estudio de casos prácticos
- MD04 – MD4. Prácticas en sala de informática
- MD05 – MD5. Seminarios
- MD06 – MD6. Ejercicios de simulación
- MD07 – MD7. Análisis de fuentes y documentos
- MD08 – MD8. Realización de trabajos en grupo
- MD09 – MD9. Realización de trabajos individuales

EVALUACIÓN (instrumentos de evaluación, criterios de evaluación y porcentaje sobre la calificación final)

EVALUACIÓN ORDINARIA

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, así como una autoevaluación razonada. Aportará el 5% de la calificación.

EVALUACIÓN EXTRAORDINARIA

La evaluación extraordinaria 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. Siendo el 50% de ponderación para cada una de ellas

EVALUACIÓN ÚNICA FINAL

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. Siendo el 50% de ponderación para cada una de ellas.

