Guía Docente

DATOS DE IDENTIFICACIÓN

Titulación:	Grado en Ingeniería Matemática
Ámbito	Ingeniería Informática y de Sistemas.
Facultad/Escuela:	Escuela Politécnica Superior
Asignatura:	Minería de Datos
Tipo:	Obligatoria
Créditos ECTS:	6
Curso:	3
Código:	4968
Periodo docente:	Sexto semestre
Materia:	Ciencia de Datos
Módulo:	Matemáticas Avanzadas y Computación
Tipo de enseñanza:	Presencial
Idioma:	Castellano
Total de horas de dedicación del alumno:	150

Equipo Docente	Correo Electrónico
Marta Salas García	marta.salas@ufv.es
Jorge Andrés Plazas Vargas	jorge.plazas@ufv.es
Rodrigo Díaz Morón	rodrigo.diaz@ufv.es

DESCRIPCIÓN DE LA ASIGNATURA

La asignatura de Minería de Datos o Data Mining proporciona al alumno las bases teóricas y prácticas para afrontar con éxito el análisis de información diversa y la extracción de información fiable a través de la exploración del contenido. Sus bases teóricas provienen de la inteligencia artificial y el aprendizaje estadístico. Además de

introducir a los alumnos a la calidad del dato, se estudian las técnicas de aprendizaje estadístico y minería de datos más habituales.

OBJETIVO

Conocer las técnicas y principios metodológicos para la extracción de información relevante a partir de distintos tipos de datos con el fin de diseñar soluciones que apoyen el proceso de toma de decisiones y permitan la realización de tareas descriptivas y predictivas.

Los fines especficos de la asignatura son:

Saber explorar un conjunto de datos para extraer la información más relevante mediante técnicas estadísticas y de visualización de la información.

Saber usar técnicas predictivas de clasificación y de regresión aplicadas a un conjunto de datos.

Saber usar técnicas descriptivas de análisis de asociación y clustering aplicadas a un conjunto de datos.

Aplicar una metodología orientada al Data Mining, como es CRISP-DM, a un problema de extracción del conocimiento.

Conocer, saber usar y saber cuándo aplicar diferentes técnicas de aprendizaje estadístico y Data Mining.

CONOCIMIENTOS PREVIOS

Se recomienda haber cursado la asignatura de "Estadística" de segundo curso.

Es deseable que el alumno tenga cierta competencia lectora en inglés. Se utilizarán algunas transparencias en inglés. No se requerirán competencias de expresión oral o escrita en inglés.

CONTENIDOS

INTRODUCCIÓN A LA **MINERÍA DE DATOS O DATA MINING**:
Tema 1. Introducción a la asignatura.
Definición de Minería de Datos. Motivación y orígenes. Tareas del Data Mining. Proceso y fases de extracción del conocimiento. Metodología CRISP-DM.
Tema 2. Preparación de datos.

Tipos de datos. Calidad de los datos. Técnicas de recopilación, almacén, limpieza y transformación de datos. Reducción de la dimensión. Técnicas de exploración y selección de datos. Estadística Descriptiva: frecuencias, percentiles, medidas de centralidad, dispersión y asimetría. Principales distribuciones de probabilidad. Selección y muestreo de datos.

Tema 3. Visualización de datos.

Histograma, diagrama de tallo y hojas, diagrama de barras, diagrama de caja, diagrama de dispersión, pie chart, diagrama de burbujas. Otras formas de visulalización: mapa coroplético (diagrama geográfico), diagrama de contornos, coordenadas paralelas, diagrama radial, diagrama treemap, diagrama de Sankey, diagrama de velocímetro (gauge), etc.

Tema 4. Análisis de asociación.

Definición del problema. Algoritmo Apriori. Extracción de reglas. Evaluación de los patrones de asociación. Manejo de atributos categóricos y continuos. Patrones secuenciales.

Tema 5. Clasificación.

Definición del problema. Clasificación basada en reglas. Árboles de decisión y algoritmos. Clasificación basada en k-Nearest-Neighbors. Probabilidad condicional, Teorema de Bayes y clasificadores. Suppor Vector Machines. Comparación y evaluación de clasificadores.

Tema 6. Predicción.

Definición del problema. Regresión lineal: coeficientes, bondad del ajuste, análisis de residuos, contraste de hipótesis. Regresión lineal multivariante. Regresión logística.

Tema 7. Clustering.

Definición del problema. Medidas de distancia. Algoritmo k-Means. Evaluación de clusters.

ACTIVIDADES FORMATIVAS

La enseñanza-aprendizaje del módulo Tecnologías Específicas atiende a una combinación armónica entre el trabajo presencial y autónomo del alumno. La asignatura de Aprendizaje Estadístico y Data Mining es eminentemente práctica. La toma de contacto con cada tema se realizará a través de metodologías expositivas por parte del profesor, que esencialmente están dirigidas a presentar brevemente los fundamentos teóricos de la asignatura. Asimismo se plantearán prácticas de aplicación de los contenidos teóricos implementadas en Python. Dichas prácticas se comenzarán a realizar en clases de laboratorio de informática con el profesor delante, pero deberán completarse de manera autónoma por parte de los alumnos fuera del aula.

Para la resolución de dudas y dificultades compartidas por diversos alumnos, el profesor podrá organizar tutorías individuales o grupales. Para favorecer la adquisición de habilidades de comunicación oral y escrita, así como del vocabulario específico de la asignatura, los alumnos se organizarán para realizar un trabajo teórico/práctico de la asignatura, del que entregarán una memoria y que tendrán que defender.

Finalmente, con el fin de facilitar al alumno el acceso a los materiales y la planificación de su trabajo, así como la comunicación con el profesor y el resto de alumnos, se empleará el Aula Virtual, que es una plataforma de aprendizaje online que ofrece diferentes recursos electrónicos para complementar, de forma muy positiva, el aprendizaje del alumno. Allí se colgarán los materiales de la asignatura, se plantearán las tareas que deben entregar los alumnos, se añadirán enlaces de interés sobre la asignatura y se habilitará un foro de comunicación entre el profesor y los alumnos.

DISTRIBUCIÓN DE LOS TIEMPOS DE TRABAJO

ACTIVIDAD PRESENCIAL	TRABAJO AUTÓNOMO/ACTIVIDAD NO PRESENCIAL
60 horas	90 horas
• Clases expositivas participativas 18h • Resolución de problemas o casos prácticos 11h • Actividades participativas grupales 5h • Seguimiento académico y actividades de evaluación 6h	• Trabajo personal y estudio autónomo 82h • Aula virtual: trabajo virtual en red, revisión y visionado de material, chats 8h

• Prácticas en laboratorio 20h	

RESULTADOS DE APRENDIZAJE

Conocer y desarrollar técnicas de aprendizaje computacional y diseñar e implementar aplicaciones y sistemas que las utilicen, incluyendo las dedicadas a extracción automática de información y conocimiento a partir de grandes volúmenes de datos.

Utilizar las técnicas matemáticas y algorítmicas necesarias para el tratamiento de datos masivos con el fin de generar conocimiento a partir de la información que ayude en la toma de decisiones.

RESULTADOS DE APRENDIZAJE ESPECIFICOS

Conocerá los fundamentos teóricos y metodológicos de las técnicas de aprendizaje y saber aplicarlos a la resolución de problemas.

Conocerá los principales métodos y algoritmos de la minería de datos.

Conocerá los conceptos subyacentes a los algoritmos probabilistas analizando en qué situaciones de la vida real es preciso acudir a este tipo de procedimientos.

SISTEMA DE EVALUACIÓN DEL APRENDIZAJE

ELEMENTOS DE EVALUACION:
[1] Examen parcial teórico-práctico escrito:
- Se realizarán a mitad de cuatrimestre y al final del curso.
- Evaluarán la adecuación de los conocimientos adquiridos por el alumno respecto de los objetivos de aprendizaje.
- Cada examen se puntuará de 0 a 10, siendo necesario obtener una nota superior o igual a 5,0 para superarlo.
- Se evaluarán el planteamiento completo, la resolución correcta, la interpretación adecuada, la justificación del resultado, la calidad de la presentación (vocabulario específico, ortografía, gramática) de cada pregunta.
[2] Prácticas en grupos:
- Están repartidas a lo largo del cuatrimestre.
- Evaluarán la aplicación de los distintos algoritmos estudiados en clase a diferentes conjuntos de datos.
- Cada práctica puntuará de 0 a 10, siendo necesario obtener una nota superior o igual a 4,0 en cada práctica.
- Se evaluarán de la memoria explicativa el planteamiento completo, la resolución correcta, la interpretación adecuada, la justificación del resultado, las fuentes de información consultadas, la calidad de la presentación (vocabulario específico, ortografía, gramática). El profesor podrá pedir que la práctica sea presentada en grupos, en cuyo caso indicará la ponderación de la presentación y de la memoria previamente y valorará el grado de conocimiento de cada alumno por la soltura en la exposición del contenido, por la adecuación de las respuestas a las preguntas, por la calidad del soporte empleado (transparencias, gráficos, etc.)
[3] Proyecto por grupos de un tema relacionado con la asignatura:
- Se entregará una memoria escrita con antelación a la defensa ante la clase de ambos miembros, y la defensa se realizará antes de la finalización de las clases.
- El proyecto puntuará de 0 a 10, siendo la memoria el 30 % y la defensa el 50 % , 10% de autoevaluación y una evaluación por grupos del 10%. La nota del trabajo será individual. Para que el proyecto contabilice en la asignatura es necesario tener una nota igual o superior a 5,0.
- Se evaluarán la dificultad del trabajo seleccionado, la claridad del planteamiento, el desarrollo, la corrección de

los resultados, la capacidad crítica de las conclusiones, las fuentes consultadas y la calidad formal (expresiones oral y escrita, material empleado para la presentación, gráficos, etc.). En la defensa, además, se valorarán individualmente las respuestas a las preguntas que haga el profesor y los compañeros).
[4] Participación activa en la asignatura:
- Participación y calidad de las intervenciones
- Presentación temprana de los trabajos
- Ampliación de los contenidos de clase
- La puntuación será entre 0 y 10.
- Este elemento no es recuperable. Para puntuar en el apartado de participación en clase, es necesario asistir al menos a un 80% de las clases.
NOTA FINAL DE LA ASIGNATURA:
El baremo de cada elemento de evaluación en la calificación final de la asignatura es el siguiente:
- Exámenes parciales: 40%
- Prácticas no guiadas: 40%
- Proyecto final: 10%
- Participación activa: 10%
Condición necesaria para aprobar:
[1] >= 5,0 [2] >= 5,0 [3] >= 5,0
Cálculo:
0,4 * [1] + 0,4 * [2] + 0,1 * [3] + 0,1 * [4]
La nota numérica de los exámenes, prácticas, trabajos y ejercicios se redondeará a una cifra decimal
RECUPERACIÓN EN CONVOCATORIA ORDINARIA:
Los alumnos que no hayan alcanzado la nota mínima en alguno de los apartados anteriores, podrán optar a su recuperación al final del cuatrimestre:
- Las notas de las partes aprobadas a lo largo del curso se guardan.
- [1] Realizarán un examen global que incluirá los dos exámenes parciales el mismo día del segundo examen parcial. Será necesario obtener una calificación superior o igual a 5,0. El primer parcial no superará la calificación tope de 6,0.
- [2] Entrega de todas la prácticas (suspensas y aprobadas) antes del examen. Es necesario que cada una de las prácticas sea superior o igual a 4,0 y que la media de todas sea superior o igual a 5. La media de las practicas no superará la calificación tope de 6,0
RECUPERACIÓN EN CONVOCATORIA EXTRAORDINARIA:
Los alumnos que no hayan alcanzado la nota mínima en alguno de los apartados anteriores tras la convocatoria ordinaria y su recuperación o deseen mejorar nota, podrán optar a su realización al final del segundo cuatrimestre:
- Las notas de las partes aprobadas a lo largo del curso se guardan.
- [1] Realizarán un examen global que incluirá los dos exámenes parciales el día señalado. Será necesario obtener una calificación superior o igual a 5,0.
- [2] Entrega de todas la prácticas (suspensas y aprobadas) antes del día señalado para el examen. Es necesario que cada una de las prácticas sea superior o igual a 4,0 y que la media de todas sea superior o igual a 5. En caso de suspender definitivamente la asignatura, no se conservan las notas para el año siguiente.
CONSUMO DE CONVOCATORIAS:
El alumno dispone de 6 convocatorias para superar esta asignatura. La Normativa de Evaluación de la UFV recoge todo lo relativo a los procesos de evaluación y consumo de convocatorias.
DISPENSA ACADÉMICA:
- Aquellos alumnos que estén exentos de la obligación de asistir a clase, bien por segunda matrícula en la asignatura o sucesivas, bien por contar con autorización expresa de la Dirección del Grado, serán evaluados por el mismo tipo de pruebas.
- Estos alumnos tienen la obligación de realizar los exámenes, las prácticas y el trabajo en los mismos plazos que el resto de sus compañeros.
- Además realizarán la defensa de forma presencial el día estipulado
- La Participación [4] se medirá con su iniciativa, participación en el foro y tener tutorías.
NORMATIVA ANTIPLAGIOS Y COPIA:
Cualquier tipo de fraude o plagio por parte del alumno en una actividad evaluable, será sancionado según se recoge en la Normativa de Convivencia de la UFV. A estos efectos, se considerará "plagio" cualquier intento de defraudar el sistema de evaluación, como copia en ejercicios, exámenes, prácticas, trabajos o cualquier otro tipo de entrega, bien de otro compañero, bien de materiales o dispositivos no autorizados, con el fin de hacer creer al profesor que son propios.

BIBLIOGRAFÍA Y OTROS RECURSOS

Básica

Jiawei Han & Micheline Kamber Data Mining: Concepts and Techniques

Complementaria

Mohammed J. Zaki & Wagner Meira, Jr. Data Mining and Machine Learning: Fundamental Concepts and Algorithms

Christopher M. Bishop Pattern Recognition and Machine Learning