GUÍA DOCENTE

DENOMINACIÓN DE LA ASIGNATURA

Denominación: **MINERÍA DE DATOS AVANZADA**
Código: 101437
Plan de estudios: **GRADO DE INGENIERÍA INFORMÁTICA** Curso: 4
Denominación del módulo al que pertenece: OPTATIVO ESPECIALIDAD COMPUTACIÓN
Materia: MINERÍA DE DATOS AVANZADA
Carácter: OPTATIVA
Duración: SEGUNDO CUATRIMESTRE
Créditos ECTS: 6.0
Horas de trabajo presencial: 60
Porcentaje de presencialidad: 40.0%
Horas de trabajo no presencial: 90
Plataforma virtual: Moodle

DATOS DEL PROFESORADO

Nombre: GARCIA PEDRAJAS, NICOLAS EMILIO (Coordinador)
Departamento: INFORMÁTICA Y ANÁLISIS NUMÉRICO
Área: CIENCIA DE LA COMPUTACIÓN E INTELIGENCIA ARTIFICIAL
Ubicación del despacho: C2
E-Mail: npedrajas@uco.es Teléfono: 957211032

Nombre: ORTIZ BOYER, DOMINGO
Departamento: INFORMÁTICA Y ANÁLISIS NUMÉRICO
Área: CIENCIA DE LA COMPUTACIÓN E INTELIGENCIA ARTIFICIAL
Ubicación del despacho: C2
E-Mail: ma1orbod@uco.es Teléfono: 957211042

REQUISITOS Y RECOMENDACIONES

Requisitos previos establecidos en el plan de estudios

No hay.

Recomendaciones

Es, a nivel práctico, imprescindible tener los conocimientos que se adquieren en la asignatura Introducción a la Minería de Datos o Introducción al Aprendizaje Automático.

COMPETENCIAS

CEC15 Conocimiento y aplicación de los principios fundamentales y técnicas básicas de los sistemas inteligentes y su aplicación práctica.

CTEC4 Capacidad para conocer los fundamentos, paradigmas y técnicas propias de los sistemas inteligentes y analizar, diseñar y construir sistemas, servicios y aplicaciones informáticas que utilicen dichas técnicas en cualquier ámbito de aplicación.

CTEC7 Capacidad para 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.

GUÍA DOCENTE

OBJETIVOS

Introducir los conceptos avanzados de la minería de datos con especial énfasis en los temas de minería de datos más relevantes en la investigación actual, como los métodos de clasificación avanzados, el escalado de algoritmos, el big data, el deep learning o los nuevos problemas.

CONTENIDOS

1. Contenidos teóricos

1.- Técnicas avanzadas de clasificación: Multiclasificadores. Agrupaciones de clasificadores. Codificación de la salida para problemas multiclase. Problemas multietiqueta, multiinstancia y multisalida. Aprendizaje semi-supervisado.
2.- Minería de datos con datos desequilibrados. Concepto de conjunto de datos desequilibrados. Métodos basados en los datos. Métodos basados en los algoritmos. Métodos mixtos.
3.- Big data: Escalado de algoritmos. Descripción del concepto de escalabilidad. Descripción de técnicas de escalado de algoritmos de minería de datos. Introducción al big data, conceptos fundamentales.
4.- Deep learning. Redes neuronales con deep learning. Otros modelos.
5.- Aplicaciones de Minería de Datos a juegos.

2. Contenidos prácticos

Implementación y uso de las técnicas explicadas en teoría en el laboratorio usando bibliotecas y herramientas apropiadas para ello.

OBJETIVOS DE DESARROLLO SOSTENIBLE RELACIONADOS CON LOS CONTENIDOS

Sin relación

METODOLOGÍA

Aclaraciones generales sobre la metodología (opcional)

Los temas se introducirán mediante lección magistral del profesor. Posteriormente los alumnos deberán exponer un artículo de investigación, a elección consensuada con el profesor, relacionado con el tema. Esta exposición formará parte de la evaluación.
En prácticas se implementarán mediante un lenguaje algunos algoritmos. Otros más complejos se usarán mediante la herramienta Weka o bibliotecas como SciKit de Python.

Adaptaciones metodológicas para alumnado a tiempo parcial y estudiantes con discapacidad y necesidades educativas especiales

Se estuadiará cada caso de forma personalizada.

Actividades presenciales

Actividad	Grupo completo	Grupo mediano	Total
Actividades de evaluación	4	2	6
Exposición grupal	10	-	10

GUÍA DOCENTE

Actividad	Grupo completo	Grupo mediano	Total
Laboratorio	-	22	22
Lección magistral	22	-	22
Total horas:	***36***	***24***	***60***

Actividades no presenciales

Actividad	Total
Búsqueda de información	10
Consultas bibliográficas	10
Ejercicios	30
Estudio	30
Problemas	10
Total horas:	***90***

MATERIAL DE TRABAJO PARA EL ALUMNO

Ejercicios y problemas
Presentaciones PowerPoint
Referencias Bibliográficas

EVALUACIÓN

Competencias	Exposición oral	Informes/memorias de prácticas	Supuesto práctico/discusión caso clínico/discusión trabajo científico
CEC15	X	X	X
CTEC4	X	X	X
CTEC7	X	X	X
Total (100%)	***40%***	***20%***	***40%***
Nota mínima (*)	***4***	***4***	***4***

(*)Nota mínima (sobre 10) necesaria para que el método de evaluación sea considerado en la calificación final de la asignatura. En todo caso, la calificación final para aprobar la asignatura debe ser igual o superior a 5,0.

GUÍA DOCENTE

Valora la asistencia en la calificación final:

No

Aclaraciones generales sobre los instrumentos de evaluación:

La evaluación se realizará mediante tres tareas:
- Exposición de artículos de investigación en clase.
- Informe de prácticas resumen de las prácticas realizadas.
- Trabajo de la asignatura voluntario de un tema relacionado con sus contenidos y consensuado con el profesor.

Los alumnos que suspendan los apartados de exposición o proyectos tendrán la oportunidad de realizar un examen final para recuperarlos.
Los alumnos que suspendan el apartado de informes de prácticas tendrán la oportunidad de recuperar con un examen final práctico.
Respecto a los instrumentos de evaluación se distribuyen según los siguientes criterios:

Bloque 1: Memorias de prácticas con un 20% en la nota final.

Bloque 2: Heteroevaluación y entrevistas sobre proyectos realizados por el alumno con un peso del 30% de la nota final.

Bloque 3: Pruebas orales a modo de exposiciones con un peso del 50% en la nota final.
Los criterio de evaluación de las las convocatorias extraordinarias de inicio de curso y fin de carrera son los mismos que las convocatorias ordinarias

Aclaraciones sobre la evaluación para el alumnado a tiempo parcial y necesidades educativas especiales:

Se estudiará cada caso de forma individualizada.

Aclaraciones sobre la evaluación de la convocatoria extraordinaria y convocatoria extraordinaria de finalización de estudios:

Se aplicarán los mismo criterios en todas las convocatorias.

Criterios de calificación para la obtención de Matrícula de Honor:

Será necesario obtener al menos 9,0 puntos. (Art. 30. del Reglamento de Régimen Académico).

BIBLIOGRAFIA

1. Bibliografía básica

Dada la naturaleza de esta asignatura la bibiografía consistirá en artículos de investigación que se pondrán a disposición del alumno.
A modo de ejemplo podemos incluir los siguientes:

1995 Solving Multiclass Learning Problems Via Error Correcting Output Codes
1999 A Survey of Methods for Scaling Up Inductive Algorithms
2000 An Experimental Comparison of Three Methods for constructing ensembles of decision trees: Bagging, Boosting and Randomization
2003 Combining classifiers: Soft computing solutions
2004 In Defense of One-Vs-All Classification
2005 Creating diversity in ensembles using artificial data
2006 Rotation Forest: A New Classifier Ensemble Method

GUÍA DOCENTE

2007 Cost-sensitive boosting for classification of imbalanced data
2007 Efficient Hierarchical Parallel Genetic Algorithms using Grid computing
2008 A memetic algorithm for evolutionary prototype selection: A scaling up approach
2008 Decision trees for hierarchical multi-label classification
2008 Large scale evolutionary optimization using cooperative coevolution
2008 Multilabel classification via calibrated label ranking
2009 A divide-and-conquer recursive approach for scaling up instance selection algorithms
2009 A Parameter-Free Classification Method for Large Scale Learning
2009 Constructing Ensembles of Classifiers by Means of Weighted Instance Selection
2010 Democratic instance selection: A linear complexity instance selection algorithm based on classifier ensemble concepts
2010 Scaling Up Support Vector Machines Using Nearest Neighbor Condensation
2011 Constructing ensembles of classifiers using supervised projection methods based on misclassified instances
2011 Decision tree induction using a fast splitting attribute selection for large datasets
2011 Efficient Distributed Genetic Algorithm for Rule extraction
2011 Evolutionary-based selection of generalized instances for imbalanced classification
2011 The effect of class imbalance on case selection for case-based classifiers: An empirical study in the context of medical decision support
2012 A Review on Ensembles for the Class Imbalance Problem: Bagging-, Boosting-, and Hybrid-Based Approaches
2012 Supervised subspace projections for constructing ensembles of classifiers
2013 Analysing the classification of imbalanced data-sets with multiple classes: Binarization techniques and ad-hoc approaches
2013 Cluster-based ensembles of classifiers
2013 EUSBoost: Enhancing ensembles for highly imbalanced data-sets by evolutionary undersampling
2014 A Review on Multi-Label Learning Algorithms
2014 A Scalable Memetic Algorithm for Simultaneous Instance and Feature Selection
2015 Classifier ensemble creation via false labelling
2015 Deep learning in neural networks: An overview
2015 Deep learning
2015 Memetic feature selection algorithm for multi-label classification
2015 Performance comparison of multi-label learning algorithms on clinical data for chronic diseases
2016 A New multi-instance multi-label learning approach for image and text classification
2016 Deep learning for visual understanding: A review
2016 Error-correcting output codes for multi-label emotion classification
2016 Recognizing emotions in text using ensemble of classifies
2016 Robust multiple-instance learning ensembles using random subspace instance selection
2017 A Divide-and-Conquer Approach for Large-scale Multi-label Learning
2017 Ant colony optimization based hierarchical multi-label classification algorithm
2017 HPSLPred: An Ensemble Multi-Label Classifier for Human Protein Subcellular Location Prediction with Imbalanced Source
2017 Measuring the class-imbalance extent of multi-class problems
2017 Multi-label classification by exploiting local positive and negative pairwise label correlation
2017 Predicting protein function via multi label supervised topic model on gene ontology
2017 Scaling-up multiobjective evolutionary clustering algorithms using stratification
2017 SCLS: Multi-label feature selection based on scalable criterion for large label set
2018 An empirical comparison on state-of-the-art multi-class imbalance learning algorithms and a new diversified ensemble learning scheme
2018 An Ensemble Oversampling Model for Class Imbalance Problem in Software Defect Prediction
2018 A study on combining dynamic selection and data preprocessing for imbalance learning
2018 Categorizing feature selection methods for multi-label classification
2018 Class Imbalance Ensemble Learning Based on the Margin Theory

GUÍA DOCENTE

2018 Class-specific extreme learning machine for handling binary class imbalance problem
2018 Hierarchical partitioning of the output space in multi-label data
2018 Learning from Ambiguously Labeled Face Images
2018 Local ensemble learning from imbalanced and noisy data for word sense disambiguation
2018 On the scalability of feature selection methods on high-dimensional data
2018 Progressive boosting for class imbalance and its application to face re-identification
2018 The Gradual Resampling Ensemble for mining imbalanced data streams with concept drift
2018 UnderBagging based reduced Kernelized weighted extreme learning machine for class imbalance learning
2018 User attribute discovery with missing labels
2019 Evaluating multi-label classifiers and recommender systems in the financial service sector
2019 Multi-label learning of part detectors for occluded pedestrian detection
2020 A-Stacking and A-Bagging: Adaptive versions of ensemble learning algorithms for spoof fingerprint detection
2020 CCE: An ensemble architecture based on coupled ANN for solving multiclass problems
2020 Ensemble learning by means of a multi-objective optimization design approach for dealing with imbalanced data sets
2020 Ensemble Selection based on Classifier Prediction Confidence

2. Bibliografía complementaria

No hay

CRITERIOS DE COORDINACIÓN

Actividades conjuntas: conferencias, seminarios, visitas...
Selección de competencias comunes

CRONOGRAMA

Periodo	Actividades de evaluación	Exposición grupal	Laboratorio	Lección magistral
1ª Quincena	0,0	0,0	4,0	6,0
2ª Quincena	0,0	2,0	4,0	4,0
3ª Quincena	0,0	0,0	4,0	6,0
4ª Quincena	0,0	2,0	4,0	4,0
5ª Quincena	0,0	2,0	4,0	2,0
6ª Quincena	0,0	2,0	2,0	0,0
7ª Quincena	6,0	2,0	0,0	0,0
Total horas:	***6,0***	***10,0***	***22,0***	***22,0***

GUÍA DOCENTE

Las estrategias metodológicas y el sistema de evaluación contempladas en esta Guía Docente serán adaptadas de acuerdo a las necesidades presentadas por estudiantes con discapacidad y necesidades educativas especiales en los casos que se requieran.