



DESCRIPCIÓN DE LA ASIGNATURA

Grado/Máster en:	Master Universitario en INGENIERÍA INFORMÁTICA por la Universidad de Málaga
Centro:	Escuela Técnica Superior de Ingeniería Informática
Asignatura:	MINERÍA DE DATOS ON-LINE (ON-LINE DATA MINING)
Código:	208
Tipo:	Optativa
Materia:	ESPECIALIDAD EN INGENIERÍA Y CIENCIA DE DATOS
Módulo:	COMPLEMENTOS EN TECNOLOGÍAS INFORMÁTICAS
Experimentalidad:	
Idioma en el que se imparte:	Español
Curso:	2
Semestre:	1
Nº Créditos	4,5
Nº Horas de dedicación del estudiante:	112,5
Nº Horas presenciales:	33,8
Tamaño del Grupo Grande:	
Tamaño del Grupo Reducido:	
Página web de la asignatura:	

EQUIPO DOCENTE

Departamento: LENGUAJES Y CIENCIAS DE LA COMPUTACIÓN

Área: LENGUAJES Y SISTEMAS INFORMÁTICOS

Nombre y Apellidos	Mail	Teléfono Laboral	Despacho	Horario Tutorías
Coordinador/a: RAFAEL MORALES BUENO	rmorales@uma.es	952131395	3.2.22 - E.T.S.I. INFORMÁTICA	
JOSE DEL CAMPO AVILA	jcampo@uma.es	951952705	3.2.34 - E.T.S.I. INFORMÁTICA	

RECOMENDACIONES Y ORIENTACIONES

Panteamiento general:

Hoy en día muchas fuentes de información, incluidas redes de sensores, mercados financieros, redes sociales o monitores médicos, generan lo que se denominan "flujos de datos" (data streams), datos que llegan de forma secuencial y a alta velocidad. Su análisis debe hacerse en tiempo real, con los datos parciales disponibles en ese momento y sin capacidad para almacenar todo el conjunto de datos. Esta asignatura presenta algoritmos y técnicas usadas en la minería de datos on-line y en el análisis de esos flujos de datos en tiempo real. Tendrá un doble enfoque de investigación y empresarial: desde el punto de vista de investigación pueden surgir nuevos métodos de procesamiento de flujos de datos; desde el punto de vista empresarial habrá sesiones de manejo de herramientas para minería de flujo de datos.

Si al problema del almacenamiento y velocidad le añadimos el problema de que los conceptos subyacentes no sean estacionarios, se hace preciso usar algoritmos online que puedan detectar los cambios y la aparición de nuevos conceptos, que es uno de los objetivos de esta asignatura.

Recomendaciones:

Haber adquirido las competencias, conocimientos y habilidades desarrolladas en la materia de Ingeniería y Ciencia de Datos (compuesta por las asignaturas Ingeniería y Ciencia de Datos I y II), que incluyen como resultados del aprendizaje:

- Conocer las bases teóricas de la Ciencia de los datos.
- Conocer, y comprender todas las fases incluidas en un Proyecto de Ingeniería de Datos.
- Conocer el proceso de ingeniería que, partiendo de grandes conjuntos de datos permite descubrir conocimiento no trivial a través de patrones y modelos con los que explicar y predecir el comportamiento de esos datos.
- Ser capaz de desarrollar un proyecto de Ingeniería de Datos siguiendo una metodología adecuada.

Y cuyos contenidos en dicha materia son:

- Estadística para el análisis de datos: Estadística descriptiva y estadística inferencial. Test y técnicas de validación.
- Modelos teóricos de aprendizaje.
- Métricas para análisis y validación de modelos.
- Metodologías de desarrollo de un sistema de Descubrimiento de Conocimiento.
- Sistemas integrados de Ingeniería de Datos.
- Preparación de datos.
- Modelado y evaluación.
- Utilización de la Ingeniería de datos en empresas. Ejemplos de casos concretos.

En todo caso la asignatura se adaptará al estado de conocimientos de los estudiantes al inicio de la misma.

CONTEXTO



El contexto académico de esta asignatura viene determinado por su entorno anterior y simultáneo.

El entorno anterior viene determinado por las asignaturas vinculadas a la materia "Ingeniería y Ciencia de Datos" (impartidas en el 1º y 2º semestre) y por otras asignaturas que se imparten en el primer curso como son "Altas Prestaciones para Datos y Ciberseguridad" (por su relación con los temas de algoritmos distribuidos) e "Industrialización y Despliegue de Sistemas IoT" (por estar vinculada a la generación de conjuntos de datos reales susceptibles de ser analizados en tiempo real).

El entorno simultáneo viene determinado por otras asignaturas que se imparten en el mismo semestre. Entre ellas:

- ¿ Especialidad de Ingeniería y Ciencia de Datos
- o Aprendizaje profundo (dotar a estos modelos con capacidades de adaptación al cambio)
- o Big Data (desplegar las herramientas online con otras tecnologías alternativas)
- o Análisis visual de datos (visualizar diferentes modelos de conocimiento y su evolución)
- ¿ Especialidad de Ciberseguridad
- o Análisis de Malware (detectar nuevos tipos de malware, detectar intrusión en sistemas)

COMPETENCIAS

1 Competencias generales y básicas.

Competencias básicas

- 1.1 CB6 - Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación
- 1.2 CB7 - Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio
- 1.3 CB8 - Que los estudiantes sean capaces de integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o limitada, incluya reflexiones sobre las responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos y juicios
- 1.4 CB9 - Que los estudiantes sepan comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades
- 1.5 CB10 - Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo

Competencias generales

- 1.4 CG4 - Capacidad para el modelado matemático, cálculo y simulación en centros tecnológicos y de ingeniería de empresa, particularmente en tareas de investigación, desarrollo e innovación en todos los ámbitos relacionados con la ingeniería en informática
- 1.8 CG8 - Capacidad para la aplicación de los conocimientos adquiridos y de resolver problemas en entornos nuevos o poco conocidos dentro de contextos más amplios y multidisciplinares, siendo capaces de integrar estos conocimientos.
- 1.9 CG9 - Capacidad para comprender y aplicar la responsabilidad ética, la legislación y la deontología profesional de la actividad de la profesión de Ingeniero en Informática.

2 Competencias específicas.

- 2.1 ET11 - Capacidad para modelar, diseñar, definir la arquitectura, implantar, gestionar, operar, administrar y mantener aplicaciones, redes, sistemas, servicios y contenidos informáticos.
- 2.2 ET12 - Capacidad de comprender y saber aplicar el funcionamiento y organización de Internet, las tecnologías y protocolos de redes de nueva generación, los modelos de componentes, software intermediario y servicios.
- 2.7 ET17 - Capacidad para comprender y poder aplicar conocimientos avanzados de computación de altas prestaciones y métodos numéricos o computacionales a problemas de ingeniería.
- 2.9 ET19 - Capacidad para aplicar métodos matemáticos, estadísticos y de inteligencia artificial para modelar, diseñar y desarrollar aplicaciones, servicios, sistemas inteligentes y sistemas basados en el conocimiento.

3 Competencias transversales.

- 3.1 CT1 - Capacidad de emprendimiento basado en la innovación, liderazgo, negociación y orientación a clientes y resultados.
- 3.2 CT2 - Capacidad para identificar estrategias, herramientas y métodos que responden a situaciones de éxito que pueden ser abordadas con los recursos disponibles.

CONTENIDOS DE LA ASIGNATURA

Introducción, cambio de concepto y técnicas de evaluación

Preprocesamiento y algoritmos de aprendizaje aproximado

Aprendizaje supervisado: clasificación y regresión



Aprendizaje no supervisado: agrupamiento

Sistemas para la minería de datos on-line e integración en sistemas propios

ACTIVIDADES FORMATIVAS

Actividades presenciales

Actividades expositivas

Lección magistral

ACTIVIDADES DE EVALUACIÓN

Actividades de evaluación presenciales

Actividades de evaluación del estudiante

Examen parcial: Prueba objetiva mediante test multirespuesta

Realización de trabajos y/o proyectos: Análisis de conjuntos de datos reales y sintéticos

Participación en clase: Exposiciones orales individuales y en grupos pequeños

RESULTADOS DE APRENDIZAJE / CRITERIOS DE EVALUACIÓN

- Conocer algoritmos y ser capaz de aplicar técnicas y herramientas para flujos de datos que se generan de forma continua y que hay que procesar on-line y en tiempo real, incluido la detección y tratamiento de cambio de tendencia/concepto (concept drift).
- Aprendizaje basado en proyectos (ABP) usando:
 - o Datos Generados artificialmente (necesarios para validar la calidad de los algoritmos en condiciones controladas), incluyendo diferentes velocidades de cambio, presencia de ruido, tiempo de respuesta al cambio, etc.;
 - o Datos reales (necesarios para hacer frente a características online y comparables a otros resultados): Covertypes, Poker-Hand, Electricity, Airlines, UseNet, Nursery;
 - o Otros conjuntos de Datos reales: Open Data (supranacionales, nacionales, locales, etc.); Competiciones (congresos): KDD Cup, ICWSM Data Challenge; Utilizados en otras asignaturas (IoT, Ing. Ciencias Datos II, etc.); Redes sociales (twitter, etc.)
- Trabajos individuales durante el curso (serán variaciones de lo hecho en clase y puede ser la base del trabajo en grupo);
- Trabajo en grupo a presentar en las últimas sesiones del curso (2 personas, aunque reconfigurable según necesidades).

PROCEDIMIENTO DE EVALUACIÓN

5% participación en clase.

70% prácticas individuales por cada bloque, basadas en variaciones a lo hecho en clase

25% trabajo en grupo (2 personas \pm 1 persona en casos justificados)

Los trabajos en grupo se presentarán a los compañeros en las últimas sesiones

Temáticas variadas (sugerencias no limitantes):

Estudio comparativo tomando los resultados de trabajos individuales

Implementar alguna variante de algoritmo (en MOA o SAMOA)

Integración en sistemas reales usando MOA o SAMOA

Aplicación a un caso real

BIBLIOGRAFÍA Y OTROS RECURSOS

Básica

- Battiti, R., & Brunato, M. (2017). The LION way. Machine Learning plus Intelligent Optimization. LIONlab, University of Trento, Italy. Retrieved from https://intelligent-optimization.org/LIONbook/lionbook_3v0.pdf
- Bifet, A., Gavaldà, R., Holmes, G., & Pfahringer, B. (2018). Machine learning for data streams: with practical examples in MOA. MIT Press. Retrieved from <https://moa.cms.waikato.ac.nz/book-html>. Disponible en la UMA: http://jabega.uma.es/record=b2260762~S4*spl
- Bifet, A. (2010). Adaptive Stream Mining: Pattern Learning and Mining from Evolving Data Streams. IOS Press. Disponible en la UMA: http://jabega.uma.es/record=b1792383~S4*spl
- Gama, J. (2010). Knowledge discovery from data streams. Chapman & Hall/CRC. Disponible en la UMA: http://jabega.uma.es/record=b1806815~S4*spl

Complementaria

- Bifet, A., Holmes, G., Kirkby, R., & Pfahringer, B. (2011). MOA Massive Online Analysis. Journal of Machine Learning Research, 11(May), 1601-1604. Retrieved from <http://www.jmlr.org/papers/volume11/bifet10a/bifet10a.pdf>
- De Francisci Morales, G., Morales, F., & Bifet, A. (2015). SAMOA: Scalable Advanced Massive Online Analysis. Journal of Machine Learning Research, 16, 149-153. Retrieved from <http://jmlr.csail.mit.edu/papers/volume16/morales15a/morales15a.pdf>



DISTRIBUCIÓN DEL TRABAJO DEL ESTUDIANTE

ACTIVIDAD FORMATIVA PRESENCIAL

Descripción	Horas	Grupo grande	Grupos reducidos
Lección magistral	33,8	<input checked="" type="checkbox"/>	<input type="checkbox"/>
TOTAL HORAS ACTIVIDAD FORMATIVA PRESENCIAL	33,8		
TOTAL HORAS ACTIVIDAD FORMATIVA NO PRESENCIAL	67,45		
TOTAL HORAS ACTIVIDAD EVALUACIÓN	11,25		
TOTAL HORAS DE TRABAJO DEL ESTUDIANTE	112,5		

