Guía Docente

DATOS DE IDENTIFICACIÓN

Titulación:	Grado en Análisis de Negocios/Business Analytics
Rama de Conocimiento:	Ciencias Sociales y Jurídicas
Facultad/Escuela:	Derecho, Empresa y Gobierno
Asignatura:	Minería de Datos II / Data Mining II
Tipo:	Optativa
Créditos ECTS:	3
Curso:	2
Código:	5348
Periodo docente:	Cuarto semestre
Materia:	Minería de Datos / Data Mining
Módulo:	Formación Disciplinar
Tipo de enseñanza:	Presencial
Idioma:	Castellano
Total de horas de dedicación del alumno:	75

Equipo Docente	Correo Electrónico
Pablo Hidalgo García	pablo.hidalgo@ufv.es

DESCRIPCIÓN DE LA ASIGNATURA

La asignatura de Minería de Datos II, continúa el trabajo iniciado de Minería de Datos I en el conocimiento y en la aplicación de técnicas avanzadas de predicción y de minería de textos, para extraer información clave. El curso ofrece formación básica en la aplicación de algoritmos avanzados mediante software R para la toma de decisiones empresariales.

OBJETIVO

Formar profesionales capaces de comprender la información contenida en los datos almacenados por las organizaciones, o la información contenida en los ficheros de texto, Internet, etc. para encontrar patrones o reglas que una vez validados puedan ser incorporados en los sistemas de análisis de información de las empresas.

Los fines especficos de la asignatura son:

Conocer técnicas avanzadas de predicción que utiliza la minería de datos para obtener unos resultados.

Comprender la información contenida en una base de datos con el uso de técnicas de Machine Learning.

Explotación de técnicas de minería de textos, con el fin de obtener valor sobre una base de datos de texto.

Conocer y saber usar el programa R para el análisis de datos y su procesamiento mediante técnicas de minería de datos.

CONOCIMIENTOS PREVIOS

Se recomienda haber cursado la asignatura de “Minería de Datos I” en el primer primer cuatrimestre de segundo curso, un nivel de comprensión de lectura medio en el idioma inglés y conocimientos básicos de programación en R.

CONTENIDOS

BLOQUE DE MODELOS AVANZADOS

1. Conceptos generales de Minería de Datos y Machine Learning: Tipos de aprendizajeAprendizaje supervisado. Regresión y clasificación.Aprendizaje no supervisado.Overfitting y underfitting.
2. Métodos basados en prototipos: K-vecinos más cercanos (KNN)
3. Modelos ensamblados. BaggingRandom forestAlgoritmos de boosting: AdaBoost

BLOQUE DE TEXT MINING

1. Introducción al text mining
2. Tratamiento básico de textos: Limpieza de texto y tokenizaciónTerm Frequency e Inverse Document Frequency (TF-IDF)Wordcloud
3. Análisis de sentimientos

ACTIVIDADES FORMATIVAS

La asignatura de Minería de Datos II tiene un enfoque eminentemente práctico.
La materia se desarrolla en dos tipos de sesiones:

- Sesiones teórico-prácticas en las que, por un lado, se llevará a cabo la exposición de los contenidos del

programa mediante el uso de clases magistrales por parte del profesor, así como con el uso de metodologías ágiles que favorecen el proceso de enseñanza-aprendizaje; y por otro lado, se realizarán ejercicios, problemas y casos sencillos relacionados con los conocimientos impartidos en la misma sesión. En estas sesiones se fomentará el uso de la metodología flipped learning o aprendizaje inverso.

- Sesiones prácticas que se desarrollarán en el aula, individualmente o en grupos de alumnos, donde se analizarán y procesarán datos obtenidos de distintas fuentes aplicando los modelos desarrollados en clase.

El alumno podrá hacer uso de tutorías tanto individuales como grupales para obtener unos mejores resultados del aprendizaje. Los trabajos en equipo exigen una supervisión continuada por parte del profesor por lo que los alumnos deberán concertar tutorías a lo largo de todo el semestre para garantizar el cumplimiento de los tiempos de realización de cada parte del proceso, así como para orientar y resolver dudas sobre el trabajo a desarrollar.

DISTRIBUCIÓN DE LOS TIEMPOS DE TRABAJO

ACTIVIDAD PRESENCIAL	TRABAJO AUTÓNOMO/ACTIVIDAD NO PRESENCIAL
30 horas	45 horas

COMPETENCIAS

Competencias básicas

Que los estudiantes hayan demostrado poseer y comprender conocimientos en un área de estudio que parte de la base de la educación secundaria general, y se suele encontrar a un nivel que, si bien se apoya en libros de texto avanzados, incluye también algunos aspectos que implican conocimientos procedentes de la vanguardia de su campo de estudio

Que los estudiantes sepan aplicar sus conocimientos a su trabajo o vocación de una forma profesional y posean las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y la resolución de problemas dentro de su área de estudio

Que los estudiantes tengan la capacidad de reunir e interpretar datos relevantes (normalmente dentro de su área de estudio) para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética

Que los estudiantes puedan transmitir información, ideas, problemas y soluciones a un público tanto especializado como no especializado

Que los estudiantes hayan desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía

Competencias generales

Capacidad de organización, sistematización y planificación en la identificación de problemas, pautas y modelos en el contexto del big data

Compromiso ético en la sociedad de la información

Capacidad para el cumplimiento de objetivos, resolución de problemas y toma de decisiones en un entorno de datos masivos tanto cuantitativos como cualitativos

Capacidad para analizar datos a gran escala procedentes de diferentes fuentes: audiovisuales, textos y numéricas

Capacidad de diseñar e implementar proyectos e informes, utilizando con naturalidad los canales digitales

Capacidad de pensamiento crítico, autocrítico, analítico y reflexivo

Capacidad de aprendizaje autónomo en la sociedad de la información

Competencias específicas

Saber identificar y resolver problemas reales de la empresa, a través del análisis avanzado de datos y de la elección de técnicas adecuadas para la toma de decisiones

Saber manejar herramientas cuantitativas e informáticas para la toma de decisiones

Conocer y saber utilizar a nivel de analista herramientas de estadística avanzada, almacén de datos, bases de datos relacionales, bases de datos no relacionales y sistemas de gestión de big data.

RESULTADOS DE APRENDIZAJE

Conoce y selecciona métodos y algoritmos de la minería de datos para identificar y resolver problemas reales de la empresa.

Utiliza con soltura el programa R en el análisis de datos para la toma de decisiones empresariales.

Conoce y aplica las técnicas de text mining aprendidas durante el curso.

Sabe extraer, comprender y analizar la información contenida en grandes volúmenes de datos.

SISTEMA DE EVALUACIÓN DEL APRENDIZAJE

El sistema de evaluación del aprendizaje será mediante evaluación continua, examen teórico-práctico, realización de ejercicios, realización y ejecución de problemas y prácticas, valoración de trabajos y de la participación en

clase. La asistencia a clases expositivas y prácticas es obligatoria para someterse a este sistema.
Todos los alumnos deberán realizar un examen final independientemente de la calificación obtenida en exámenes parciales, evaluación continua, pruebas, etc. Para hacer media, la calificación obtenida en dicha prueba final debe ser superior o igual a 5. La nota de los trabajos realizados durante el desarrollo de la asignatura se mantendrá para la convocatoria extraordinaria. Estos trabajos han de ser presentados en el tiempo indicado por los profesores; aquellos trabajos entregados fuera del plazo establecido computarán como 0.
En ningún caso la superación de la asignatura podrá reducirse a la aprobación de un examen. Todas las pruebas susceptibles de evaluación estarán supeditadas a lo establecido en la normativa de evaluación de la Universidad Francisco de Vitoria.
Las conductas de plagio, así como el uso de medios ilegítimos en las pruebas de evaluación, serán sancionados conforme a los establecido en la Normativa de Evaluación y la Normativa de Convivencia de la universidad.
Los profesores se reservan el derecho de llamar al alumno para la defensa oral de trabajos y/o exámenes para comprobar su autoría.

SISTEMA DE EVALUACIÓN PARA ALUMNOS DE PRIMERA MATRÍCULA

- Actividades diarias y ejercicios: 15 %
- Trabajos individuales y grupales: 15%
- Trabajo transdisciplinar: 15%
- Asistencia y participación en las actividades presenciales en el aula: 5 %
- Examen final: 50%

ALUMNOS CON DISPENSA ACADÉMICA

- Trabajos presentados a requerimiento del profesor: 30%
- Examen final (será el mismo que los alumnos de asistencia normal): 70%

ALUMNOS DE SEGUNDA Y SUCESIVAS MATRÍCULAS

- En este caso los alumnos podrán optar entre cualquiera de los dos sistemas anteriores siempre y cuando se haya pedido autorización al PEC y a los profesores de la asignatura los 5 primeros días de clase por escrito. En caso contrario, se aplicará por defecto el sistema de evaluación de primera matrícula.

CONVOCATORIA EXTRAORDINARIA

- Se aplicará el mismo criterio que en los casos anteriores.

El tiempo fijado para los trabajos prácticos está programado para ejecutarse en ese tiempo, por lo que las fechas de entrega se han de cumplir. El calendario de entrega de estos trabajos se proporcionará al comienzo del curso.

NOTA SOBRE PLAGIO

Los alumnos están obligados a observar las reglas elementales y básicas sobre autenticidad y originalidad en toda actividad formativa o prueba de evaluación. Cuando un alumno disponga o se valga de medios ilegítimos en la celebración de una prueba de evaluación, incurra en plagio, o se atribuya indebidamente la autoría de trabajos académicos requeridos para la evaluación será sancionado conforme a lo establecido en la Normativa de Evaluación y Convivencia de la universidad

BIBLIOGRAFÍA Y OTROS RECURSOS

Básica

Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani An Introduction to Statistical Learning 2021

Robert E. Schapire, Yoav Freund Boosting 2014

Julia Silge, David Robinson Text mining with R 2018