



Universidad
Internacional
de Valencia

FICHA INFORMATIVA

ASIGNATURA: Captura y preparación de datos (tipología y fuentes de datos)

Título: *Captura y preparación de datos (tipología y fuentes de datos)*

Carácter: *Obligatoria*

Créditos: *6 ECTS*

Código:

Índice

1. Organización general.....	3
1.1. Datos de la asignatura	3
1.2. Introducción a la asignatura	3
1.3. Competencias y resultados de aprendizaje	3
2. Contenidos	4
3. Metodología	4
4. Actividades formativas	5
5. Evaluación	8
5.1. Sistema de evaluación	8
5.2. Sistema de calificación	8

1. Organización general

1.1. Datos de la asignatura

ASIGNATURA	<i>Captura y preparación de datos (tipología y fuentes de datos)</i> 6 ECTS
Carácter	Obligatoria
Curso	Segundo
Semestre	Cuarto
Idioma en que se imparte	Castellano
Requisitos previos	No existen
Dedicación al estudio por ECTS	25 horas

1.2. Introducción a la asignatura

En esta asignatura se analizan técnicas de captura y agregación de datos provenientes de múltiples fuentes en un mismo espacio de almacenamiento, manteniendo la coherencia entre ellos, así como de los procesos de limpieza y preparación de datos mediante técnicas de ingeniería de características, incluyendo eliminación de outliers o valores extremos, reducción de dimensionalidad, transformaciones de los datos, web scraping, etc.; se usan herramientas para tratamiento de datos y la extracción de información/conocimiento, como Apache Kafka, Flume, y el lenguaje de programación Python.

1.3. Competencias y resultados de aprendizaje

C05 - Desarrollar modelos de simulación estadísticos y matemáticos para soportar la toma de decisiones en las empresas e instituciones.

C11 - Utilizar los procesos teóricos y técnicas aplicadas para el tratamiento de datos provenientes de fuentes estructuradas, no estructuradas y multimodales y su integración de cara a la inferencia de nuevo conocimiento.

C12 - Comprender las características, funcionalidades, estructura y los modelos de datos para la implementación eficiente de aplicaciones basadas en bases de datos relacionales y NoSQL.

C13 - Elaborar informes que resuman, interpreten y contrasten de forma crítica los resultados obtenidos utilizando las herramientas de análisis y visualización de grandes volúmenes de datos.

C14 - Utilizar aplicaciones informáticas de aprendizaje automático, estadística avanzada, visualización gráfica y optimización para la resolución de problemas en el ámbito de la Ciencia de Datos.

C15 - Aplicar los aspectos éticos, legales y normativos relacionados con el tratamiento y requisitos de privacidad de los datos y la explotación del conocimiento obtenido.

C16 - Definir los requerimientos de los sistemas informáticos para dimensionar la infraestructura de soporte que garantice la adquisición, el almacenamiento y el procesamiento distribuido de los datos.

CC10 - Conocer los retos éticos, legales y sociales que plantean los avances de la Ciencia de Datos y la Inteligencia Artificial, en los ámbitos nacional, europeo e internacional.

CC8 - Dominar los conceptos fundamentales de la ciencia de los datos, que incluyen las fases del ciclo de vida del dato, los roles de

un científico de datos, las tecnologías big data y la computación en la nube. TIPO: Conocimientos o contenidos

CC9 - Conocer las arquitecturas físicas y las herramientas de software de sistemas informáticos que permitan el procesamiento de

altas prestaciones y almacenamiento de grandes volúmenes de datos.

H01 - Habilidad de pensamiento analítico, algorítmico y de abstracción para la identificación y análisis de soluciones adecuadas a problemas complejos de análisis de grandes volúmenes de datos.

H02 - Habilidad para desarrollar actividades en el ámbito de la Ciencia de Datos y la Inteligencia Artificial asumiendo un compromiso responsable social, ético, sostenible y profesional.

H06 - Desarrollar un pensamiento crítico que contribuya al discernimiento eficaz en el desarrollo de actividades profesionales propias de un científico de datos.

H08 - Habilidad para favorecer una visión holística de soluciones eficientes, creativas e inteligentes a los problemas basados en datos para la toma de decisiones en las empresas.

H09 - Desarrollar la intuición y entendimiento de los productos para realizar análisis cuantitativos y predecir el comportamiento del sistema.

2. Contenidos

1. Tipos de fuentes y tipos de datos en entornos big data.
2. Técnicas para enlazar datos de distintas fuentes.
3. Pre-procesamiento de datos, limpieza de datos, transformación de datos, selección de variables, selección de instancias, preparación de datos.
4. Búsqueda y análisis de información textual, modelos generales (e.g., LSI, LDA).
5. Búsqueda y análisis en estructuras enlazadas, captura de datos de la web, crawling and scraping, redes sociales, web semántica.
6. Algoritmos para datos masivos, flujos de datos (streaming), muestras, tratamiento de datos sensibles.
7. Ejercicios prácticos con software específico, como Apache Kafka, Flume, y el lenguaje de programación Python..

3. Metodología

La metodología de la Universidad Internacional de Valencia (VIU) se caracteriza por una apuesta decidida en un modelo de carácter e-presencial. Así, siguiendo lo estipulado en el calendario de

actividades docentes del Título, se impartirán en directo un conjunto de sesiones, que, además, quedarán grabadas para su posterior visionado por parte de aquellos estudiantes que lo necesiten. En todo caso, se recomienda acudir, en la medida de lo posible, a dichas sesiones, facilitando así el intercambio de experiencias y dudas con el docente.

En lo que se refiere a las metodologías específicas de enseñanza-aprendizaje, serán aplicadas por el docente en función de los contenidos de la asignatura y de las necesidades pedagógicas de los estudiantes. De manera general, se impartirán contenidos teóricos y, en el ámbito de las clases prácticas se podrá realizar la resolución de problemas, el estudio de casos y/o la simulación.

Por otro lado, la Universidad y sus docentes ofrecen un acompañamiento continuo al estudiante, poniendo a su disposición foros de dudas y tutorías para resolver las consultas de carácter académico que el estudiante pueda tener. Es importante señalar que resulta fundamental el trabajo autónomo del estudiante para lograr una adecuada consecución de los objetivos formativos previstos para la asignatura.

4. Actividades formativas

Durante el desarrollo de cada una de las asignaturas se programan una serie de actividades de aprendizaje que ayudan a los estudiantes a consolidar los conocimientos trabajados.

Actividades formativas		
Actividad	Horas	Presencialidad
Clases expositivas	10	0%
Clases prácticas	10	0%
Tutorías	16	0%
Trabajo autónomo	112	0%
Prueba objetiva final	2	100%

A continuación, se relacionan las actividades que forman parte de la asignatura:

1. Actividades de carácter teórico

Se trata de un conjunto de actividades guiadas por el profesor de la asignatura destinadas a la adquisición por parte de los estudiantes de los contenidos teóricos de la misma. Estas actividades, diseñadas de manera integral, se complementan entre sí y están directamente relacionadas con los materiales teóricos que se ponen a disposición del estudiante (manual, SCORM y material complementario). Estas actividades se desglosan en las siguientes categorías:

- a. Clases expositivas

- b. Sesiones con expertos en el aula
- c. Observación y evaluación de recursos didácticos audiovisuales
- d. Estudio y seguimiento de material interactivo

2. Actividades de carácter práctico

Se trata de un conjunto de actividades guiadas y supervisadas por el profesor de la asignatura vinculadas con la adquisición por parte de los estudiantes de los resultados de aprendizaje y competencias de carácter más práctico. Estas actividades, diseñadas con visión de conjunto, están relacionadas entre sí para ofrecer al estudiante una formación completa e integral.

3. Tutorías

Se trata de sesiones, tanto de carácter síncrono como asíncrono (e-mail), individuales o colectivas, en las que el profesor comparte información sobre el progreso académico del estudiante y en las que se resuelven dudas y se dan orientaciones específicas ante dificultades concretas en el desarrollo de la asignatura.

4. Trabajo autónomo

Se trata de un conjunto de actividades que el estudiante desarrolla autónomamente y que están enfocadas a lograr un aprendizaje significativo y a superar la evaluación de la asignatura. La realización de estas actividades es indispensable para adquirir las competencias y se encuentran entroncadas en el aprendizaje autónomo que consagra la actual ordenación de enseñanzas universitarias. Esta actividad, por su definición, tiene carácter asíncrono.

5. Prueba objetiva final

Como parte de la evaluación de cada una de las asignaturas (a excepción de las prácticas y el Trabajo fin de título), se realiza una prueba (examen final). Esta prueba se realiza en tiempo real (con los medios de control antifraude especificados) y tiene como objetivo evidenciar el nivel de adquisición de conocimientos y desarrollo de competencias por parte de los estudiantes. Esta actividad, por su definición, tiene carácter síncrono.

Metodologías docentes	
LMA	Lección magistral
LMP	Lección magistral participativa
SEM	Seminarios y conferencias on-line
ECA	Estudio de casos
RPR	Resolución de problemas

RBI	Revisión bibliográfica
SIM	Simulación

TCO	Trabajo Cooperativo
DPR	Diseño de proyectos
SEG	Seguimiento

5. Evaluación

5.1. Sistema de evaluación

El Modelo de Evaluación de estudiantes en la Universidad se sustenta en los principios del Espacio Europeo de Educación Superior (EEES), y está adaptado a la estructura de formación virtual propia de esta Universidad. De este modo, se dirige a la evaluación de competencias.

Sistema de Evaluación	Ponderación Mínima	Ponderación Máxima
Portafolio*	40 %	60 %
Prueba final*	40 %	60 %

***Es requisito indispensable para superar la asignatura aprobar cada apartado (portafolio y prueba final) con un mínimo de 5 para ponderar las calificaciones.**

Los enunciados y especificaciones propias de las distintas actividades serán aportados por el docente, a través del Campus Virtual, a lo largo de la impartición de la asignatura.

Atendiendo a la Normativa de Evaluación de la Universidad, se tendrá en cuenta que la utilización de **contenido de autoría ajena** al propio estudiante debe ser citada adecuadamente en los trabajos entregados. Los casos de plagio serán sancionados con suspenso (0) de la actividad en la que se detecte. Asimismo, el uso de **medios fraudulentos durante las pruebas de evaluación** implicará un suspenso (0) y podrá implicar la apertura de un expediente disciplinario.

5.2. Sistema de calificación

La calificación de la asignatura se establecerá en los siguientes cálculos y términos:

Nivel de aprendizaje	Calificación numérica	Calificación cualitativa
Muy competente	9,0 - 10	Sobresaliente
Competente	7,0 - 8,9	Notable

Aceptable	5,0 -6,9	Aprobado
Aún no competente	0,0 -4,9	Suspenso

Sin detrimento de lo anterior, el estudiante dispondrá de una **rúbrica simplificada** en el aula que mostrará los aspectos que valorará el docente, como así también los **niveles de desempeño que tendrá en cuenta para calificar las actividades vinculadas a cada resultado de aprendizaje.**

La mención de «**Matrícula de Honor**» podrá ser otorgada a estudiantes que hayan obtenido una calificación igual o superior a 9.0. Su número no podrá exceder del cinco por ciento de los estudiantes matriculados en una materia en el correspondiente curso académico, salvo que el número de estudiantes matriculados sea inferior a 20, en cuyo caso se podrá conceder una sola «Matrícula de Honor».