



Guía de Asignatura

ASIGNATURA: Big Data

Título: Grado en Ingeniería Informática

Materia: Big data

Créditos: 6 ECTS

Código: XXGIIN

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.....	4
2.	Contenidos	5
3.	Metodología.....	6
4.	Actividades formativas.....	7
5.	Evaluación	10
5.1.	Sistema de evaluación	10
5.2.	Sistema de calificación	10
6.	Bibliografía	12
6.1.	Bibliografía de referencia	12

1. Organización general

1.1. Datos de la asignatura

TITULACIÓN	Grado en Ingeniería Informática
ASIGNATURA	Big Data
CÓDIGO - NOMBRE ASIGNATURA	XXGIIN_Big Data
Carácter	Optativo
Curso	Cuarto
Cuatrimestre	Segundo
Idioma en que se imparte	Castellano
Requisitos previos	No existen
Dedicación al estudio por ECTS	25 horas

1.2. Introducción a la asignatura

La asignatura de Big Data aborda la creciente importancia y complejidad de gestionar grandes volúmenes de datos en la era actual. Comenzando con una introducción a los desafíos y oportunidades que presenta el Big Data, se exploran las herramientas y técnicas necesarias para su procesamiento eficiente.

En cuanto a los entornos de programación, se profundiza en Apache Hadoop como una plataforma líder para la computación distribuida, permitiendo el procesamiento de datos a gran escala a través de múltiples nodos. Se enseña cómo manejar datos en estos entornos, incluyendo la programación distribuida en nodos y la configuración de sistemas de gestión de procesos.

Para abordar la manipulación de datos a gran escala, se introduce Apache Spark, una tecnología que facilita el procesamiento rápido de grandes volúmenes de datos en clústeres de computadoras. Se cubren aspectos clave como la ingestión de datos, la limpieza e integración de los mismos, así como la ingeniería de características para preparar los datos para su análisis.

En el ámbito del machine learning, se exploran técnicas eficientes tipo Random Fourier Features, Orthogonal Fourier Features y Nystrom, centrándose en su

aplicación a los kernel methods para el aprendizaje escalable. Se utiliza PySpark para implementar algoritmos de aprendizaje supervisado, incluyendo regresiones y clasificaciones, abordando aspectos de escalabilidad y gestión del ciclo de vida en el proceso de desarrollo de modelos.

En resumen, la asignatura proporciona una visión integral de cómo abordar los desafíos del Big Data, desde la ingestión y limpieza de datos hasta la implementación de algoritmos de machine learning escalables, utilizando herramientas como Apache Hadoop, Apache Spark y PySpark.

1.3. Competencias y resultados de aprendizaje

COMPETENCIAS BÁSICAS

C.B.1.- 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.

C.B.2.- 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.

C.B.3.- 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.

C.B.4.- Que los estudiantes puedan transmitir información, ideas, problemas y soluciones a un público tanto especializado como no especializado.

C.B.5. - Que los estudiantes hayan desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía.

COMPETENCIAS ESPECÍFICAS DE LA MENCIÓN

C.E.M.C.6.- Desarrollar software capaz de escalar con una gran cantidad de datos.

2. Contenidos

- Introducción al big data y a la problemática de los datos en la era actual.
- Técnicas eficientes de machine learning para la escalabilidad de algoritmos tipo Random Fourier Features, Orthogonal Fourier Features, Nystrom y su aplicación a los kernel methods.
- Entornos de programación big data tipo hadoop apache para computación distribuida.
- Manejo de big data en entornos de programación.
- Programación distribuida en nodos.
- Configuración de sistemas de gestión de procesos tipo Slurm.

3. Metodología

La modalidad de enseñanza propuesta para el presente título guarda consonancia con la Metodología General de la Universidad Internacional de Valencia, aprobada por el Consejo de Gobierno Académico de la Universidad y de aplicación en todos sus títulos.

Este modelo, que vertebra el conjunto de procesos de enseñanza y aprendizaje de la institución, combina la naturaleza síncrona (mismo tiempo-diferente espacio) y asíncrona (diferente tiempo -diferente espacio) de los entornos virtuales de aprendizaje, siempre en el contexto de la modalidad virtual.

El elemento síncrono se materializa en sesiones de diferente tipo (clases expositivas y prácticas, tutorías, seminarios y actividades de diferente índole durante las clases online) donde el profesor y el estudiante comparten un espacio virtual y un tiempo determinado que el estudiante conoce con antelación.

Las actividades síncronas forman parte de las actividades formativas necesarias para el desarrollo de la asignatura y, además, quedan grabadas y alojadas para su posterior visualización.

Por otro lado, estas sesiones no solamente proporcionan espacios de encuentro entre estudiante y profesor, sino que permiten fomentar el aprendizaje colaborativo, al generarse grupos de trabajo entre los estudiantes en las propias sesiones.

Los elementos asíncronos del modelo se desarrollan a través del Campus Virtual, que contiene las aulas virtuales de cada asignatura, donde se encuentran los recursos y contenidos necesarios para el desarrollo de actividades asíncronas, así como para la interacción y comunicación con los profesores y con el resto de departamentos de la Universidad.

4. Actividades formativas

La metodología VIU, basada en la modalidad virtual, se concreta en una serie de actividades formativas y metodologías docentes que articulan el trabajo del estudiante y la docencia impartida por los profesores.

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 en cada una de las asignaturas. A continuación, listamos las actividades genéricas que pueden formar parte de cada asignatura, dependiendo de las competencias a desarrollar en los estudiantes en cada asignatura.

1. Clases virtuales síncronas

Constituyen el conjunto de acciones formativas que ponen en contacto al estudiante con el profesor, con otros expertos y con compañeros de la misma asignatura en el mismo momento temporal a través de herramientas virtuales. Las actividades recurrentes (por ejemplo, las clases) se programan en el calendario académico y las que son ocasionales (por ejemplo, sesiones con expertos externos) se avisan mediante el tablón de anuncios del campus. Estas actividades se desglosan en las siguientes categorías:

a. Clases expositivas: El profesor expone a los estudiantes los fundamentos teóricos de la asignatura.

b. Clases prácticas: El profesor desarrolla junto con los estudiantes actividades prácticas que se basan en los fundamentos vistos en las clases expositivas. En términos generales, su desarrollo consta de las siguientes fases, pudiéndose adaptar en función de las necesidades docentes:

I. La primera fase se desarrolla en la sala principal de la videoconferencia, donde el profesor plantea la actividad.

II. A continuación, divide a los estudiantes en grupos de trabajo a través de las salas colaborativas y se comienza con la actividad. En esta fase el profesor va entrando en cada sala colaborativa rotando los grupos para resolver dudas, dirigir el trabajo o dar el feedback oportuno. Los estudiantes también tienen posibilidad de consultar al profesor en el momento que consideren necesario.

III. La tercera fase también se desarrolla en la sala principal y tiene como objetivo mostrar el ejercicio o explicar con ejemplos los resultados obtenidos. Por último, se ponen en común las conclusiones de la actividad realizada.

No obstante, el profesor puede utilizar otras metodologías activas y/o herramientas de trabajo colaborativo en estas clases.

c. Seminarios: En estas sesiones un experto externo a la Universidad acude a presentar algún contenido teórico-práctico directamente vinculado con el temario de la asignatura. Estas sesiones permiten acercar al estudiante a la realidad de la disciplina en términos no sólo profesionales, sino también académicos. Todas estas sesiones están vinculadas a contenidos de las asignaturas y del programa educativo.

2. Actividades asíncronas supervisadas

Se trata de un conjunto de actividades supervisadas por el profesor de la asignatura vinculadas con la adquisición por parte de los estudiantes de los resultados de aprendizaje y el desarrollo de sus competencias. Estas actividades, diseñadas con visión de conjunto, están relacionadas entre sí para ofrecer al estudiante una formación completa e integral. Esta categoría se desglosa en el siguiente conjunto de actividades:

a. Actividades y trabajos prácticos: se trata de un conjunto de actividades prácticas realizadas por el estudiante por indicación del profesor que permiten al estudiante adquirir las competencias del título, especialmente aquellas de carácter práctico. Estas actividades, entre otras, pueden ser de la siguiente naturaleza: actividades vinculadas a las clases prácticas (resúmenes, mapas conceptuales, one minute paper, resolución de problemas, análisis reflexivos, generación de contenido multimedia, exposiciones de trabajos, test de autoevaluación, participación en foros, entre otros). Estas actividades serán seleccionadas por el profesor en función de las necesidades docentes. Posteriormente, estas actividades son revisadas por el profesor, que traslada un feedback al estudiante sobre las mismas, pudiendo formar parte de la evaluación continua de la asignatura.

b. Actividades guiadas con recursos didácticos audiovisuales e interactivos: se trata de un conjunto de actividades en las que el estudiante revisa o emplea recursos didácticos (bibliografía, videos, recursos interactivos) bajo las indicaciones realizadas previamente por el profesor; con el objetivo de profundizar en los contenidos abordados en las sesiones teóricas y prácticas. Estas sesiones permiten la reflexión o práctica por parte del estudiante, y pueden complementarse a través de la puesta en común en clases síncronas o con la realización de actividades y trabajos prácticos. Posteriormente, estas actividades son revisadas por el profesor, que traslada un feedback al estudiante sobre las mismas, pudiendo formar parte de la evaluación continua de la asignatura.

3. Tutorías

En esta actividad se engloban las sesiones virtuales de carácter síncrono y las comunicaciones por correo electrónico o campus virtual destinadas a la tutorización de los estudiantes. En ellas, el profesor comparte información sobre el progreso del trabajo del estudiante a partir de las evidencias recogidas, se resuelven dudas y se dan orientaciones específicas ante dificultades concretas en el desarrollo de la asignatura. Pueden ser individuales o colectivas, según las necesidades de los estudiantes y el carácter de las dudas y orientaciones planteadas. Tal y como se ha indicado, se realizan a través de videoconferencia y e-mail.

Se computan una serie de horas estimadas, pues, aunque existen sesiones comunes para todos los estudiantes, éstos posteriormente pueden solicitar al docente tantas tutorías como estimen necesarias.

Dado el carácter mixto de esta actividad formativa, se computa un porcentaje de sincronía estimado del 30%.

4. Estudio autónomo

En esta actividad el estudiante consulta, analiza y estudia los manuales, bibliografía y recursos propios de la asignatura de forma autónoma a fin de lograr un aprendizaje significativo y superar la evaluación de la asignatura de la asignatura. Esta actividad es indispensable para adquirir las competencias del título, apoyándose en el aprendizaje autónomo como complemento a las clases y actividades supervisadas.

5. Examen 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 o 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. Los exámenes o pruebas de evaluación final se realizan en las fechas y horas programadas con antelación y con los sistemas de vigilancia online (proctoring) de la universidad.

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

***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 9.5. 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».

6. Bibliografía

6.1. Bibliografía de referencia

- Nudurupati, O'Reilly for Higher Education (Firm), and an O. M. Company. Safari, (2024, Accessed: Mar. 04, 2024). Essential PySpark for Scalable Data Analytics, p. 322, [Online]. Available: <https://www.packtpub.com/product/essential-pyspark-for-scalabledata-analytics/9781800568877>
- Drabas, T., Lee, D. and Karau, H. (2016, Accessed: Mar. 04, 2024). Learning PySpark : build data-intensive applications locally and deploy at scale using the combined powers of Python and Spark 2.0. [Online]. Available: <https://www.oreilly.com/library/view/learning-pyspark/9781786463708/>
- Feng, W. (2018) Learning Apache Spark with Python.