

Guía Docente: Fundamentos de los Sistemas de Big Data

| DATOS GENERALES | |
|------------------------------------|--|
| Facultad | Facultad de Ciencias y Tecnología |
| Titulación | Máster en Análisis Inteligente de Datos (Big Data) |
| Plan de estudios | 2018 |
| Materia | Sistemas Big Data y análisis estadístico |
| Carácter | Obligatorio |
| Período de impartición | Primer Trimestre |
| Curso | Primero |
| Nivel/Ciclo | Máster |
| Créditos ECTS | 6 |
| Lengua en la que se imparte | Castellano |
| Prerrequisitos | No se prevén requisitos previos; por tanto, los requisitos serán los propios del título. |

| DATOS DEL PROFESORADO | | | |
|-------------------------------|--|---------------------------|-----------------------------------|
| Profesor Responsable | Eugenio Marinetto Carrillo | Correo electrónico | eugenio.marinetto@ui1.es |
| Área | | Facultad | Facultad de Ciencias y Tecnología |
| Perfil Profesional 2.0 | LinkedIn Google Scholar | | |

CONTEXTUALIZACIÓN Y JUSTIFICACIÓN DE LA ASIGNATURA

| | |
|--|---|
| Asignaturas de la materia | <ul style="list-style-type: none"> • Fundamentos de los Sistemas de Big Data • Análisis estadístico y visualización de datos |
| Contexto y sentido de la asignatura en la titulación y perfil profesional | <p>El objetivo de esta asignatura es dar una introducción transversal al alumno en materia del tratamiento y obtención de valor de los datos. A lo largo del plan docente, el alumno aprenderá los aspectos más relevantes a tener en cuenta para el análisis de datos que ya realizan hoy día grandes compañías y se está convirtiendo, de facto, en uno de los activos más importantes de las mismas.</p> <p>En la actualidad, el gran volumen de datos tratados impone requisitos tecnológicos específicos para poder abordar los problemas de procesamiento y almacenamiento de los mismos. Dichas tecnologías se engloban en lo que se conoce como Big Data. El alumno estudiará dichas necesidades y requisitos, así como las principales tecnologías tanto de almacenamiento, procesamiento y análisis que se utilizan en la actualidad para la gestión de los datos de forma masiva.</p> <p>Con el cambio de legislación vigente en Europa y la inclusión de GDPR, muchas de estas tecnologías y sistemas de tratamiento masivo de datos se han visto envueltas en nuevos requisitos legales. A lo largo de la asignatura, el alumno estudiará dicha legislación y su aplicación directa en diferentes casos de uso reales que podrán encontrarse en su futuro puesto laboral.</p> <p>Otro de los aspectos vanguardistas de las tecnologías que envuelven al mundo de los datos es el del Cloud. Proveedores de servicios como AWS, Microsoft Azure y Google Cloud Platform ofrecen la posibilidad a cualquier usuario de construir sistemas Big Data de forma sencilla, transparente y económica. La aparición de las tecnologías y servicios gestionados del Cloud democratiza el empleo de dichas tecnologías de forma global y facilita la implantación de dichos sistemas. El alumno aprenderá de forma práctica a comprender y analizar dichas arquitecturas y sistemas con el objetivo de introducirlo en el mundo Cloud que cada vez más impera en la industria.</p> <p>Finalmente, no podemos olvidarnos de la inteligencia artificial, tan de moda en los últimos años por las capacidades predictivas que ofrece y que aportan muchísimo valor a las empresas en su desarrollo y estrategia. El alumno aprenderá el proceso de creación y desarrollo de modelos de Machine Learning y se introducirá en los conceptos clave del Deep Learning y Redes Neuronales de forma práctica con el objetivo de prepararlos para un mundo laboral que cada vez demanda más profesionales de estas áreas.</p> |

COMPETENCIAS QUE ADQUIERE EL ESTUDIANTE Y RESULTADOS DE APRENDIZAJE

| | |
|---|---|
| Competencias de la asignatura | <ul style="list-style-type: none"> • 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. • CB7: Saber aplicar los conocimientos adquiridos y ser capaz de resolver problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio. • CB8: Ser capaz 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. • CB9: Ser capaz de transmitir sus conclusiones, y los conocimientos y fundamentos que las sustentan, a públicos especializados y no especializados de un modo claro y sin ambigüedades. • 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. • CG1: Conocer la legislación en materia de datos personales, privacidad y derechos fundamentales de las personas, interoperabilidad, propiedad intelectual y reutilización de datos. • CG6: Ser capaz de integrarse en equipos de trabajo o investigación multidisciplinares de manera eficaz y colaborativa. • CE01: Comprender, aplicar y analizar arquitecturas y técnicas propias de bigdata para el análisis de datos estáticos y dinámicos, estructurados y no estructurados. • CE05: Diseñar, desarrollar y probar soluciones bigdata adaptadas para la captación, almacenamiento y tratamiento de grandes volúmenes de datos procedentes de diferentes contextos. • CE06: Comprender y utilizar técnicas avanzadas de visualización de datos y de experiencia de usuario para el diseño e implementación de interfaces adaptadas al usuario en los procesos de análisis de bigdata en distintos contextos. |
| Resultados de aprendizaje de la asignatura | <ul style="list-style-type: none"> • Comprender la legislación en materia de datos personales y privacidad, reconociendo las posibilidades de reutilización de los datos en función de su tipología específica. • Conocer los conceptos y métodos de modelado de datos más utilizados en big data, y manejar sistemas de gestión de BD NoSQL. • Analizar e identificar la forma más óptima de almacenamiento de los datos para un acceso eficiente que permita alta disponibilidad y tolerancia a fallos, distinguiendo diferentes tipos de datos (multimedia, texto, estructurado, no estructurado). • Conocer los conceptos fundamentales del aprendizaje automático, sus tipos y modelos. Comprender las agrupaciones de clustering y la formación y utilidades de las redes neuronales. |

PROGRAMACION DE CONTENIDOS

| | |
|---|--|
| Breve descripción de la asignatura | <p>En esta asignatura se verá, entre otros:</p> <ul style="list-style-type: none"> • Legislación sobre datos y reutilización. • Gestión de sistemas de BDNóSQL. • Fuentes de datos. Obtención, integración y análisis. • Análisis de datos multimedia. • Reconocimiento de patrones. Redes neuronales |
|---|--|

Contenidos**Unidad didáctica 1. El nuevo petróleo**

- Requisitos y gestión del dato
- Datos
 - Estructura
 - Origen y tratamiento
 - Temporalidad
 - Volumen
- Protección de datos
 - GDPR
- Gobernabilidad del dato
 - Seguridad y privacidad del dato
 - Responsabilidades y procesos
- Anexo: Características de los datos de hoy

Unidad didáctica 2. Big Data

- Definición de Big Data
 - Volumen
 - Velocidad
 - Variedad
 - Veracidad
 - Valor
- Tecnologías Big Data
 - Necesidades y requisitos
 - Arquitectura
 - Almacenamiento
 - Procesamiento
 - Analítica y Business Intelligence
 - Inteligencia artificial
 - Productivización y buenas prácticas

Unidad didáctica 3. Tratamiento de datos

- Extracción
 - Fuentes de datos
 - Formatos
 - Streaming
 - Gobernabilidad
- Transformación
 - EDA: Análisis Exploratorio de Datos
 - Limpieza
 - Transformaciones comunes
 - Nulos
 - Calidad del dato
- Carga
 - Agregación
 - Persistencia
 - El valor del dato

Unidad didáctica 4. Big Data en el Cloud

- Introducción al cloud
 - Definición
 - Servicios gestionados
 - Modelo de responsabilidad compartida con los proveedores cloud
 - Microservicios

- Servicios cloud
 - AWS S3 Simple Storage Service
 - AWS Athena
 - AWS Lambda
 - AWS SNS Simple Notification Service
 - AWS SQS Simple Queue Service
 - AWS Dynamo DB
 - AWS Neptune
- Fundamentos de arquitectura cloud
 - Ejemplo de arquitectura big data

Unidad didáctica 5. Introducción al Machine Learning

- Introducción
- Tipos de modelos ML
 - Regresión y clasificación
 - Supervisados y no supervisados
- Componentes de un modelo ML
 - Ingesta
 - Feature engineering
 - Selección del modelo
 - Entrenamiento
 - Selección de hiperparámetros
 - Productivización
- Modelo
 - Regresión lineal
 - Regresión lineal con características polinómicas
 - Regresión logística
 - K-means

Unidad didáctica 6. Machine Learning Avanzado y Deep Learning

- Machine Learning avanzado
 - Train – Test
 - Validación Cruzada (cross-validation)
 - Overfitting y Underfitting
 - Modelos en cascada
- Redes Neuronales
 - Modelado del cerebro humano
 - La neurona
 - Capas
- Workflow para el desarrollo de un modelo Big Data
 - Envejecimiento de los modelos

METODOLOGÍA

Actividades formativas

- **Cuestionarios evaluables:** Cada conjunto de dos unidades didácticas se propondrá un cuestionario de preguntas tipo test, similares a las que el alumno se encontrará en el examen final, abordando los conocimientos abordados en dichas unidades didácticas. Éstos computarán para la nota final; y además en ellos, los alumnos y alumnas valorarán la comprensión de los contenidos de las unidades didácticas.
- **Actividades de descubrimiento inducido (Estudio de caso).** Actividades en las que el alumno podrá llevar a cabo un aprendizaje contextualizado trabajando, en el Aula Virtual y de manera colaborativa, una situación real o simulada que le permitirá realizar un primer acercamiento a los diferentes temas de estudio.
- **Actividades de Interacción y colaboración (Foros-Debates de apoyo al caso y a la lección).** Actividades en las que se discutirá y argumentará acerca de diferentes temas relacionados con las asignaturas de cada materia y que servirán para guiar el proceso de descubrimiento inducido.
- **Actividades de aplicación práctica.** Incluye la resolución de problemas, elaboración de proyectos y actividades similares que permitan aplicar los aspectos conceptuales, procedimentales y actitudinales trabajados en otras partes de las asignaturas.
- **Trabajos de investigación:** Los alumnos deberán realizar una documentación e investigación acerca de un tema propuesto por el docente. Como resultado del mismo, se deberá realizar una memoria que contenga los resultados de la investigación desarrollada.
- **Tutorías.** Permiten la interacción directa entre docente y alumno para la resolución de dudas y el asesoramiento individualizado sobre distintos aspectos de la asignatura.
- **Actividades de autoevaluación.** A lo largo de cada unidad didáctica se incluyen actividades para que el alumno pueda evaluar la asimilación de los contenidos teóricos.

EVALUACIÓN

Sistema evaluativo

En caso de que la situación sanitaria impida la realización presencial de los exámenes con todas las garantías, la Universidad Isabel I celebrará dichas pruebas en modalidad online. Para la realización de dichos exámenes, la universidad incorporará la herramienta de proctoring a nuestra plataforma tecnopedagógica, con el objetivo de garantizar los procesos de autenticación del alumno, como el control del entorno durante el desarrollo de las pruebas de evaluación. A su vez, la Universidad Isabel I pondrá a disposición del alumnado una Unidad de Exámenes Online específica para ofrecer apoyo técnico durante todo el proceso y así solventar todas las incidencias que se puedan presentar.

El sistema de evaluación se basará en una selección de las pruebas de evaluación más adecuadas para el tipo de competencias que se trabajen. El sistema de calificaciones estará acorde con la legislación vigente (*Real Decreto 1125/2003, de 5 de septiembre, por el que se establece el sistema europeo de créditos y el sistema de calificaciones en las titulaciones universitarias de carácter oficial y de validez en todo el territorio nacional*).

El sistema de evaluación de la Universidad Isabel I queda configurado de la siguiente manera:

Sistema de evaluación convocatoria ordinaria

Opción 1. Evaluación continua

Los estudiantes que opten por esta vía de evaluación deberán realizar el **seguimiento de la evaluación continua (EC)** y podrán obtener hasta un **60 %** de la calificación final a través de las actividades que se plantean en la evaluación continua.

Además, deberán realizar un **examen final presencial (EX)** que supondrá el **40 %** restante. Esta prueba tiene una parte dedicada al control de la identidad de los estudiantes que consiste en la verificación del trabajo realizado durante la evaluación continua y otra parte en la que realizan diferentes pruebas teórico-prácticas para evaluar las competencias previstas en cada asignatura.

Para la aplicación de los porcentajes correspondientes, el estudiante debe haber obtenido una nota mínima de un 4 en cada una de las partes de las que consta el sistema de evaluación continua.

Se considerará que el estudiante supera la asignatura en la convocatoria ordinaria por el sistema de evaluación continua, siempre y cuando al aplicar los porcentajes correspondientes se alcance una calificación mínima de un 5.

Opción 2. Prueba de evaluación de competencias

Los estudiantes que opten por esta vía de evaluación deberán realizar una **prueba de evaluación de competencias (PEC)** y un **examen final presencial (EX)**.

La **PEC** se propone como una prueba que el docente plantea con el objetivo de evaluar en qué medida el estudiante adquiere las competencias definidas en su asignatura. Dicha prueba podrá ser de diversa tipología, ajustándose a las características de la asignatura y garantizando la evaluación de los resultados de aprendizaje definidos. Esta prueba supone el 50 % de la calificación final.

El **examen final presencial**, supondrá el **50 %** de la calificación final. Esta prueba tiene una parte dedicada al control de la identidad de los estudiantes que consiste en la verificación del seguimiento de las actividades formativas desarrolladas en el aula virtual y otra parte en la que realizan diferentes pruebas teórico-prácticas para evaluar las competencias previstas en cada asignatura.

Al igual que con el sistema de evaluación anterior, para la aplicación de los porcentajes correspondientes el estudiante debe haber obtenido una puntuación mínima de un 4 en cada una de las partes de las que consta la opción de prueba de evaluación de competencias.

Se considerará que el estudiante supera la asignatura en la convocatoria ordinaria por el sistema de la prueba de evaluación de competencias siempre y cuando al aplicar los porcentajes correspondientes se alcance una calificación mínima de un 5.

Características de los exámenes

Los exámenes constarán de 30 ítems compuestos por un enunciado y cuatro opciones de respuesta, de las cuales solo una será la correcta. Tendrán una duración de 90 minutos y la calificación resultará de otorgar 1 punto a cada respuesta correcta, descontar 0,33 puntos por cada respuesta incorrecta y no puntuar las no contestadas. Después, con el resultado total, se establece una relación de proporcionalidad en una escala de 10.

Sistema de evaluación convocatoria extraordinaria

Todos los estudiantes, independientemente de la opción seleccionada, que no superen las pruebas evaluativas en la convocatoria ordinaria tendrán derecho a una convocatoria extraordinaria.

La convocatoria extraordinaria completa consistirá en la realización de una **prueba de evaluación de competencias** que supondrá el **50 %** de la calificación final y un **examen final presencial** cuya calificación será el **50 %** de la calificación final.

Para la aplicación de los porcentajes correspondientes, el estudiante debe haber obtenido una nota mínima de un 4 en cada una de las partes de las que consta el sistema de evaluación de la convocatoria extraordinaria.

Los estudiantes que hayan suspendido todas las pruebas evaluativas en convocatoria ordinaria (evaluación continua o prueba de evaluación de competencias y examen final) o no se hayan presentado deberán realizar la convocatoria extraordinaria completa, como se recoge en el párrafo anterior.

En caso de que hayan alcanzado una puntuación mínima de un 4 en alguna de las pruebas evaluativas de la convocatoria ordinaria (evaluación continua o prueba de evaluación de competencias y examen final), se considerará su calificación para la convocatoria extraordinaria, debiendo el estudiante presentarse a la prueba que no haya alcanzado dicha puntuación o que no haya realizado.

En el caso de que el alumno obtenga una puntuación que oscile entre el 4 y el 4,9 en las dos partes de que se compone la convocatoria ordinaria (EC o PEC y examen), solo se considerará para la convocatoria extraordinaria la nota obtenida en la evaluación continua o prueba de evaluación de competencias ordinaria (en función del sistema de evaluación elegido), debiendo el alumno realizar el examen extraordinario para poder superar la asignatura.

Al igual que en la convocatoria ordinaria, se entenderá que el alumno ha superado la materia en convocatoria extraordinaria si, aplicando los porcentajes correspondientes, se alcanza una calificación mínima de un 5.

BIBLIOGRAFÍA Y OTROS RECURSOS

Bibliografía básica

Erl, T., Khattak, W., & Buhler, P. (2016). *Big Data Fundamentals: Concepts, Drivers & Techniques*. Prentice Hall.

Introducción a los conceptos fundamentales del Big Data desde un punto de vista pragmático. Teoría y terminología fundamentales de Big Data con numerosos ejemplos y diagramas sencillos. Aplicación de Big Data desde un punto de vista práctico y de negocio.

Géron, A. (2017). *Hands-On Machine Learning with Scikit-Learn and TensorFlow*. O'Reilly.

Este libro hace un repaso detallado y didáctico sobre técnicas de Machine Learning desde un punto de vista práctico, con ejercicios de programación en Python. Temas como Scikit-Learn, TensorFlow 2.0 y Keras son librerías muy utilizadas en el desarrollo de soluciones de Inteligencia Artificial que son detalladamente explicadas desde cero a lo largo de este libro.

Bibliografía complementaria

- Jackson, K. L., & Goessling, S. (2018). *Architecting Cloud Computing Solutions*. Packt Publishing.
- MongoDB: The Definitive Guide, 2nd Edition. Powerful and Scalable Data Storage. Kristina Chodorow. O'Reilly. 2013.
- Mastering Apache Cassandra – Second Edition. Nishant Neeraj. Packt Publishing. 2015.
- Beginning Neo4j. Chris Kemper. Apress. 2016.
- Redis: Learning Redis. Design efficient web and business solutions with Redis. Vinoo Das. Packt Publishing. 2015.

Otros recursos

- Curso Deep Learning Gratuito: <https://github.com/iamtrask/Grokking-Deep-Learning>
- Curso Machine Learning para principiantes: <https://github.com/humphd/have-fun-with-machine-learning>
- Blog muy interesante con conceptos avanzados de ML y DL: <https://colah.github.io/>
- Web de aprendizaje de diferentes temas Cloud y Software como Microservicios y APIs: <https://learn.freecodecamp.org/>
- Web con laboratorios gratuitos y formación en Microsoft Azure (Cloud): <https://docs.microsoft.com/en-us/learn/>
- Resumen de Servicios Gestionados AWS: <https://interactive.linuxacademy.com/diagrams/TheOrionPapers.html>
- Curso de Machine Learning Gratuito: <https://mlcourse.ai/>
- Laboratorios gratuitos y de pago en Google Cloud y AWS: <https://run.qwiklabs.com/>
- AI School de Microsoft: <https://aischool.microsoft.com/en-us/machine-learning/learning-paths/python-data-science-notebook-by-jake-vanderplas/python-data-science-handbook>
- Base de conocimientos de Databricks: <https://databricks.gitbooks.io/databricks-spark-knowledge-base/content/index.html>
- White Papers de AWS: Documentación acerca de todos los servicios AWS <https://aws.amazon.com/es/whitepapers/>
- Cassandra: <http://cassandra.apache.org/>. Acceso 25 de julio de 2019.
- Neo4j: <https://neo4j.com/>. Acceso 25 de julio de 2019.
- MongoDB: <https://docs.mongodb.com/>. Acceso 25 de julio de 2019.

