

## Datos básicos

**Número de créditos:** 69,00 ECTS

**Preinscripción:** A partir del 15/07/2019

**Matrícula:** A partir del 01/09/2019

**Impartición:** Del 09/10/2019 al 30/11/2020

**Precio (euros):** 2.095,00 (tasas incluidas)

**Pago fraccionado:** Sí

**Modalidad:** Presencial

**Prácticas en empresa/institución:** Sí (extracurriculares)

**Procedimientos de Evaluación:** Asistencia, Trabajos

### Dirección

**Unidad Organizadora:**

Departamento de Ciencias de la Computación e Inteligencia Artificial

**Director de los estudios:**

D. Joaquín Borrego Díaz

### Requisitos

Titulados en Matemáticas, Estadística, Computación, Física, Económicas y Empresariales, Ingenieros, así como profesionales ligados a puestos TIC o a la gestión y explotación de datos en empresas y administraciones públicas (en posesión de la titulación universitaria requerida para el acceso a los estudios de posgrado de acuerdo a la normativa de Enseñanzas Propias).

### Objetivos

- Habilitar al alumno en las competencias necesarias para el tratamiento computacional y estadístico de datos, utilizando técnicas computacionales y de Inteligencia Artificial.
- Formar a titulados en el análisis científico de datos.
- Implantar unos estudios de postgrado que capaciten al egresado como científico de datos dentro de la empresa.
- Facilitar una formación aplicada en Data Science a titulados superiores.
- Capacitar a titulados en el manejo científico de soluciones para Big Data.
- Proveer al empresariado e instituciones de la región de especialistas cualificados en la explotación de datos y extracción de conocimiento de los mismos.
- Actualizar las competencias y conocimientos de profesionales en Estadística o Computación que trabajan en el sector público o privado.

### Comisión Académica

D. Joaquín Borrego Díaz. Universidad de Sevilla - Ciencias de la Computación e Inteligencia Artificial

D. José A. Alonso Jiménez. Universidad de Sevilla - Ciencias de la Computación e Inteligencia Artificial

D. José Antonio Guerrero Durán. - Servicio de Documentación Clínica del Hospital Virgen del Rocío

D. Juan Luis Moreno Rebollo. Universidad de Sevilla - Estadística e Investigación Operativa

D. Juan Manuel Muñoz Pichardo. Universidad de Sevilla - Estadística e Investigación Operativa

### Profesorado

Puede consultar la lista completa del profesorado en:

<https://cfp.us.es/cursos/mu/data-science-y-big-data/5893/>

## Máster Propio

2019-2020

- **Data Science y Big Data**  
(IV edición)



### Información

Teléfono: 954 55 78 79

Web: <http://masterds.es/>

Twitter: <https://twitter.com/mscdatascience>

Email: [jborrego@us.es](mailto:jborrego@us.es)



# Asignaturas del Curso

## Módulo/Asignatura 1. Diseño y Gestión de Proyectos en Data Science I

Número de créditos: 3,00 ECTS

**Contenido:** • Introducción a la Ciencia del Dato y Big Data. • Experimentación e Investigación. Reproducibilidad y documentación. • Proyectos básicos de aplicación de técnicas de Ciencia del Dato y Big data. • Etapas de un proyecto. • Prospectiva Tecnológica. • Cuestiones éticas y legales asociadas a cada etapa de un proyecto. • Legislación de protección de datos que afectan al proyecto en Data Science.

**Fechas de inicio-fin:** 10/10/2019 - 15/03/2020

**Horario:** Miércoles y jueves, en horario de tarde.

## Módulo/Asignatura 2. Diseño y Gestión de Proyectos en Data Science II

Número de créditos: 6,00 ECTS

**Contenido:** • Técnicas de anonimización. Aplicaciones a datos médicos. • Calidad del dato. • Metodologías para incrementar la calidad de la información. • Enriquecimiento: Open Data, Linked Data, Semantic Linked Data, etc. • Aprendizaje relacional. • Ciencia del Dato y empresa. • Problemas empresariales y soluciones basadas en Data Science. • Modelos de negocio basados en datos. • Casos de estudio.

**Fechas de inicio-fin:** 28/03/2020 - 14/07/2020

**Horario:** Miércoles, jueves y viernes, en horario de tarde.

## Módulo/Asignatura 3. Introducción a la Programación en Python y a los Paradigmas para Datos

Número de créditos: 4,00 ECTS

**Contenido:** • Fundamentos de Python. • Tipos de datos: números, cadenas, tuplas, listas, diccionarios, conjuntos. • Definiciones por comprensión, estructuras de control, etc. • Definición de funciones. • Gestión de errores. • Arquitecturas y paradigmas para los datos (Introducción). • Tipos de datos: Estructurados y desestructurados. • Obtención de datos. • Almacenamiento y consulta (BDR). • Biblioteca estándar Python y otras bibliotecas.

**Fechas de inicio-fin:** 11/10/2019 - 12/12/2019

**Horario:** Jueves y viernes, en horario de tarde.

## Módulo/Asignatura 4. Fundamentos de Estadística y Programación en R

Número de créditos: 4,00 ECTS

**Contenido:** • Fundamentos de Estadística: Probabilidad e inferencia estadística. Análisis Exploratorio de Datos. Modelo de Regresión Lineal Múltiple. • Entorno de programación estadística R: Tipos de objetos. Elementos de Programación. Gestión de datos. Análisis exploratorio de datos: visualización.

**Fechas de inicio-fin:** 11/10/2019 - 23/11/2019

**Horario:** Miércoles, jueves y viernes, en horario de tarde.

## Módulo/Asignatura 5. Aprendizaje Estadístico y Modelización

Número de créditos: 6,00 ECTS

**Contenido:** • Conceptos básicos en Aprendizaje estadístico. • Modelo lineal generalizado. • Modelos de Regresión no lineales. • Modelos de Regresión no paramétricos. • Regresión y clasificación mediante KNN.

**Fechas de inicio-fin:** 28/11/2019 - 30/01/2020

**Horario:** Miércoles, jueves y viernes, en horario de tarde.

## Módulo/Asignatura 6. Arquitecturas y Paradigmas para Ciencia del Dato

Número de créditos: 4,00 ECTS

**Contenido:** • Introducción a los datos (estructurados y desestructurados). • Introducción y repaso sobre obtención de datos (APIs, RDF, logs, transacciones, textos, etc.). • Tratamiento y limpieza de datos: Expresiones regulares, Herramientas ETL. • Almacenamiento: BD relacionales versus BD NoSQL. • Bases de datos NoSQL. • Procesamiento en paralelo y escalabilidad. • Introducción a los ecosistemas Big Data.

**Fechas de inicio-fin:** 14/12/2019 - 11/01/2020

**Horario:** Miércoles, jueves y viernes, en horario de tarde.

## Módulo/Asignatura 7. Arquitecturas y Paradigmas para Big Data

Número de créditos: 5,00 ECTS

**Contenido:** • Introducción a los ecosistemas Big Data. • Almacenamiento y procesamiento distribuido. • El ecosistema Hadoop. • Spark. • Integración con R y Python.

**Fechas de inicio-fin:** 25/04/2020 - 31/05/2020

**Horario:** Miércoles, jueves y viernes, en horario de tarde.

## Módulo/Asignatura 8. Machine Learning I

Número de créditos: 5,00 ECTS

**Contenido:** • Técnicas de Reducción de la dimensionalidad. • Análisis de Conglomerados. • Técnicas de clasificación probabilística. • Support Vector Machines (SVM). • Árboles de clasificación y regresión.

**Fechas de inicio-fin:** 01/02/2020 - 14/03/2020

**Horario:** Miércoles, jueves y viernes, en horario de tarde.

## Módulo/Asignatura 9. Machine Learning II

Número de créditos: 5,00 ECTS

**Contenido:** • Redes neuronales. • Sistemas de ranking. • Técnicas de combinación de modelos. • Herramientas estadísticas para la Bioinformática. • Selección de atributos.

**Fechas de inicio-fin:** 15/03/2020 - 03/05/2020

**Horario:** Miércoles, jueves y viernes, en horario de tarde.

## Módulo/Asignatura 10. Metaheurística y Deep Learning

Número de créditos: 4,00 ECTS

**Contenido:** • Métodos de soluciones individuales. • Métodos basados en poblaciones. • Redes convolucionales. • Redes recurrentes.

**Fechas de inicio-fin:** 06/06/2020 - 13/07/2020

**Horario:** Martes y miércoles, en horario de tarde.

## Módulo/Asignatura 11. Inteligencia Colectiva y Sistemas de Recomendación

Número de créditos: 3,00 ECTS

**Contenido:** • Inteligencia Colectiva y flujos de datos: Web Social. Social Media. Metodologías "crowd". • Redes: Conceptos, topología y comportamiento extremal: Propagación. Resiliencia en redes. Computación bajo reputación y confianza. • Análisis de la semántica emergente: Análisis de opinión y sentimientos. • Sistemas de recomendación (Item based, User Based). • Aplicaciones.

**Fechas de inicio-fin:** 18/01/2020 - 06/02/2020

**Horario:** Miércoles, jueves y viernes, en horario de tarde.

## Módulo/Asignatura 12. Modelos para Datos Temporales y Espaciales

Número de créditos: 3,00 ECTS

**Contenido:** • Técnicas básicas. • Modelos ARIMA. • Modelos de series temporales no lineales: ARCH-GARCH. • Análisis espectral. • Georeferenciación estadística. • Técnicas de predicción espacial.

**Fechas de inicio-fin:** 16/05/2020 - 05/06/2020

**Horario:** Miércoles, jueves y viernes, en horario de tarde.

## Módulo/Asignatura 13. Data Science & Business Intelligence

Número de créditos: 3,00 ECTS

**Contenido:** • Introducción a la inteligencia de negocio. • Aplicaciones. • Data Smart y Data Warehouse. • Bases de datos OLAP, OLTP. • Plataformas de inteligencia de negocio. • Herramientas de minería de datos integradas en las plataformas. • Uso de herramientas de BI. • Diseño de cuadro de mandos.

**Fechas de inicio-fin:** 18/06/2020 - 12/07/2020

**Horario:** Miércoles, jueves y viernes, en horario de tarde.

## Módulo/Asignatura 14. Procesamiento del Lenguaje Natural para la Ciencia del Dato

Número de créditos: 4,00 ECTS

**Contenido:** • **Módulo I: Fundamentos del Procesamiento del Lenguaje Natural:** • Introducción: de la Lingüística Computacional a la Ingeniería del Lenguaje Natural. • Niveles y técnicas para el estudio y tratamiento automático de los lenguajes naturales. • Nociones y técnicas básicas: Expresiones regulares, Autómatas y Gramáticas. • Recursos lingüísticos. • **Módulo II: Lenguaje natural, datos y conocimiento:** • Reconocimiento automático de la voz: Redes neuronales y otras tecnologías aplicadas al nivel acústico. Análisis del entrenamiento, adaptación y evaluación. • Análisis léxico-morfológico: corpus lingüísticos y técnicas relacionadas con la ciencia del dato. Corrección ortográfica, lematización, tokenización. • Análisis gramatical: sintaxis y semántica. Modelos basados en conocimiento y modelos estadísticos. • Gestión del diálogo: Autómatas. Sistemas basados en agentes. El enfoque basado en el estado de información. Las técnicas POMDP (procesos de decisión de Markov parcialmente observables) como soporte al aprendizaje automático en gestión del diálogo. • Traducción automática: Técnicas y enfoques. Paradigmas basados en reglas, basados en corpus y enfoques híbridos. • Aplicaciones comerciales e industriales: Speech Analytics, Content Analytics, Text Mining, Information Retrieval and Extraction, Question Answering, Sentiment Analysis, Entity Names Recognition, etc. • NLTK: Librería Python para lenguaje natural. • Manipulación de corpus. • Normalización de textos: segmentación, tokenización, lematización. • Tagging. • Clasificación de textos. • Parsing. • Análisis semántico: gramáticas de estructuras de rasgos. • Representación del conocimiento. • StanfordCore NLP. • Sistema de reconocimiento de entidades. • Tagging. • Classifier. • Information Extraction. • Estudio de aplicaciones comerciales e industriales. • Ejemplos como Watson Analytics y otras plataformas que integran Big Data y sistemas basados en lenguaje natural en distintas áreas.

**Fechas de inicio-fin:** 07/02/2020 - 03/04/2020

**Horario:** Miércoles, jueves y viernes, en horario de tarde.

## Módulo/Asignatura 15. Visualización de Datos

Número de créditos: 2,00 ECTS

**Contenido:** • Herramientas, librerías y metodologías para la visualización de datos.

**Fechas de inicio-fin:** 22/02/2020 - 13/03/2020

**Horario:** Miércoles, jueves y viernes, en horario de tarde.

## Módulo/Asignatura 16. Visualización de la Información

Número de créditos: 2,00 ECTS

**Contenido:** • Arquitectura de la visualización de la Información. • Estructura visual: diseño e implementación. • Infografías y economía visual. • Presentación de informes y Reports.

**Fechas de inicio-fin:** 03/06/2020 - 14/06/2020

**Horario:** Miércoles, jueves y viernes, en horario de tarde.

## Módulo/Asignatura 17. Trabajo Fin de Máster

Número de créditos: 6,00 ECTS

**Contenido:** • Desarrollo del Trabajo fin de Máster, orientado a la aplicación de los conocimientos y competencias adquiridas en los restantes módulos. Los tutores serán tanto académicos como empresariales.

**Fechas de inicio-fin:** 01/03/2020 - 14/10/2020