

**DATA SCIENCE Y BIG DATA (II EDICIÓN)**

**Datos básicos del Curso**

Curso Académico	2021 - 2022
Nombre del Curso	Data Science y Big Data (II Edición)
Tipo de Curso	Máster Propio
Número de créditos	69,00 ECTS

**Dirección**

Unidad organizadora	Departamento de Ciencias de la Computación e Inteligencia Artificial
Director de los estudios	D Joaquín Borrego Díaz

**Requisitos**

**Requisitos específicos de admisión a los estudios**  
 Titulados en Matemáticas, Estadística, Computación, Ingenieros, Física, Económicas y Empresariales, 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).

**Requisitos académicos para la obtención del Título o Diploma**  
 Superar satisfactoriamente los 69 créditos ECTS de que se compone el Máster

**Criterios de selección de alumnos**

Criterio I. Expediente académico de la titulación. (Máximo 2 puntos). Criterio II. Perfil académico (Máximo 3 puntos).

Grupo A. Matemáticas, Estadística, Computación, Informática. Grupo B. Física, Económicas y Empresariales, Ingenierías  
 Grupo C. Otras titulaciones de grado o licenciaturas.

Criterio III. Situación laboral (Máximo 3 puntos).

Grupo A. Parado o trabajando como profesionales ligados a puestos TIC o a la gestión y explotación de datos en empresas y administraciones públicas  
 Grupo B. Otras situaciones laborales.

Criterio IV. Experiencia profesional (Máximo 1 punto).

Grupo A. Profesionales ligados a puestos TIC o a la gestión y explotación de datos en empresas y administraciones públicas  
 Grupo B. Otros perfiles profesionales.

Criterio IV. Carta personal de presentación y/o entrevista personal (Máximo 1 punto).

Carta personal dirigida al director de los estudios (o entrevista personal) que recoja los motivos por los que desea realizar el máster y la adecuación de su perfil a los criterios de admisión arriba recogidos.

<b>Preinscripción</b>	Fecha de inicio	15/07/2021
	Fecha de fin	31/08/2021
<b>Datos de Matriculación</b>	Fecha de inicio	01/09/2021
	Fecha de fin	11/09/2021
	Precio (euros)	2.344,00 (tasas incluidas)
	Pago fraccionado	Sí
<b>Impartición</b>	Fecha de inicio	11/10/2021
	Fecha de fin	17/07/2022
	Modalidad	A distancia
	Idioma impartición	Español
	Plataforma virtual	Plataforma Virtual US
	Prácticas en empresa/institución	No
<b>Información</b>	Teléfono	954557879
	Web	<a href="https://masterds.es">https://masterds.es</a>
	Facebook	
	Twitter	
	Email	<a href="mailto:jborrego@us.es">jborrego@us.es</a>

## DATA SCIENCE Y BIG DATA (II EDICIÓN)

### Objetivos del Curso

- \* 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.

### Competencias Generales

CG01. Capacidad para aplicar los conocimientos adquiridos y para resolver problemas en entornos nuevos o poco conocidos dentro de contextos más amplios y multidisciplinares, siendo capaces de integrar estos conocimientos.

CG02. Capacidad 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.

CG03. Capacidad de comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades

CG04. 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.

CG05. Los estudiantes serán capaces de realizar un análisis crítico, evaluación y síntesis de ideas nuevas y complejas.

CG06. Los estudiantes deben ser capaces de comunicarse con sus colegas, con la comunidad con la que comparte intereses profesionales en su conjunto y con la sociedad en general acerca de sus áreas de conocimiento.

CG07. Los estudiantes serán capaces de fomentar, en contextos académicos y profesionales, el avance tecnológico, social o cultural dentro de una sociedad basada en el conocimiento.

CG08. Los estudiantes serán capaces de analizar textos del área en otras lenguas relevantes en el ámbito científico.

CG09. Los estudiantes serán capaces de evaluar la calidad de nuevos métodos de creación y gestión de conocimiento.

CG10. Capacidad para formular opiniones globales y razonadas sobre proyectos de ámbito creativo-tecnológico.

### Procedimientos de Evaluación

Pruebas, Trabajos, Participación en actividades

### Comisión Académica

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

D. José Antonio Guerrero Durán. Institución no universitaria - Servicio de Documentación Clínica del Hospital Virgen del Rocío

D. Juan Galán Páez. Universidad de Sevilla - Ciencias de la Computación e Inteligencia Artificial

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

D. Rafael Blanquero Bravo. Universidad de Sevilla - Estadística e Investigación Operativa

### Profesorado

D. Pedro Almagro Blanco. Universidad de Sevilla - Ciencias de la Computación e Inteligencia Artificial

D. Gonzalo Antonio Aranda Corral. Universidad de Huelva- Tecnologías de la Información

D<sup>a</sup>. Inmaculada Barranco Chamorro. Universidad de Sevilla - Estadística e Investigación Operativa

D. Antonio Beato Moreno. Universidad de Sevilla - Estadística e Investigación Operativa

D. Rafael Blanquero Bravo. Universidad de Sevilla - Estadística e Investigación Operativa

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

D<sup>a</sup>. M<sup>a</sup> Dolores Cubiles de la Vega. Universidad de Sevilla - Estadística e Investigación Operativa

D. Javier de la Rosa Pérez. - Research Scientist (Biblioteca Nacional de Noruega)

D. Jaime de Miguel Rodríguez. - Arquitecto

D. Juan Galán Páez. Universidad de Sevilla - Ciencias de la Computación e Inteligencia Artificial

D. David Gálvez Ruiz. Universidad de Sevilla - Estadística e Investigación Operativa

D. Felipe García de Pesquera Gago. - Centro de Estudios Universitarios "Cardenal Spínola"

D. Jesús Giráldez Crú. Universidad de Granada- Ciencias de la Computación e Inteligencia Artificial

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

D<sup>a</sup>. María Dolores Jiménez Gamero. Universidad de Sevilla - Estadística e Investigación Operativa

D. Antonio Jiménez Mavillard. University Western Ontario - Canadá- Big Data

D. Pedro Luis Luque Calvo. Universidad de Sevilla - Estadística e Investigación Operativa

D. Francisco J. Martín Mateos. Universidad de Sevilla - Ciencias de la Computación e Inteligencia Artificial

D. Miguel Ángel Martínez del Amor. Universidad de Sevilla - Ciencias de la Computación e Inteligencia Artificial

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

D. Rafael Pino Mejías. Universidad de Sevilla - Estadística e Investigación Operativa

D. José Luis Pino Mejías. Universidad de Sevilla - Estadística e Investigación Operativa

D. José Francisco Quesada Moreno. Universidad de Sevilla - Ciencias de la Computación e Inteligencia Artificial

D<sup>a</sup>. Ana María Rojas Mendoza. - Consejo Superior de Investigaciones Científicas - Centro Andaluz de Biología del Desarrollo

D. Álvaro Romero Jiménez. Universidad de Sevilla - Ciencias de la Computación e Inteligencia Artificial

D. José L. Ruiz Reina. Universidad de Sevilla - Ciencias de la Computación e Inteligencia Artificial

D. Fernando Sancho Caparrini. Universidad de Sevilla - Ciencias de la Computación e Inteligencia Artificial

D. David Solís Martín. Universidad de Sevilla - Ciencias de la Computación e Inteligencia Artificial

### Módulos/Asignaturas del Curso

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

Número de créditos: 9,00 ECTS

Modalidad de impartición: A distancia

Contenido: \*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

\*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: 11/10/2021 - 24/06/2022

Horario: Estudios a distancia, Módulo/Asignatura sin horario

### Módulo/Asignatura 2. Introducción a la Programación en Python y a los Paradigmas para Datos

Número de créditos: 4,00 ECTS

Modalidad de impartición: A distancia

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: 11/10/2021 - 25/11/2021

Horario: Estudios a distancia, Módulo/Asignatura sin horario

### Módulo/Asignatura 3. Fundamentos de Estadística y Programación en R

Número de créditos: 4,00 ECTS

Modalidad de impartición: A distancia

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: 11/10/2021 - 25/11/2021

Horario: Estudios a distancia, Módulo/Asignatura sin horario

#### Módulo/Asignatura 4. Aprendizaje Estadístico y Modelización

Número de créditos: 6,00 ECTS

Modalidad de impartición: A distancia

Contenido: \* Concepto 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: 26/11/2021 - 13/01/2022

Horario: Estudios a distancia, Módulo/Asignatura sin horario

#### Módulo/Asignatura 5. Arquitecturas y Paradigmas para Ciencia del Dato

Número de créditos: 4,00 ECTS

Modalidad de impartición: A distancia

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: 13/11/2021 - 11/12/2021

Horario: Estudios a distancia, Módulo/Asignatura sin horario

#### Módulo/Asignatura 6. Arquitecturas y Paradigmas para Big Data

Número de créditos: 5,00 ECTS

Modalidad de impartición: A distancia

Contenido: \* Introducción a los ecosistemas Big Data

- \* Almacenamiento y procesamiento distribuido \* El ecosistema Hadoop
- \* Spark
- \* Integración con R y Python.

Fechas de inicio-fin: 24/03/2022 - 05/05/2022

Horario: Estudios a distancia, Módulo/Asignatura sin horario

### Módulo/Asignatura 7. Machine Learning I

Número de créditos: 5,00 ECTS

Modalidad de impartición: A distancia

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: 14/01/2022 - 17/02/2022

Horario: Estudios a distancia, Módulo/Asignatura sin horario

### Módulo/Asignatura 8. Machine Learning II

Número de créditos: 5,00 ECTS

Modalidad de impartición: A distancia

Contenido: \* Redes neuronales

- \* Sistemas de ranking
- \* Técnicas de combinación de modelos.
- \* Herramientas estadísticas para la Bioinformática \* Selección de atributos

Fechas de inicio-fin: 18/02/2022 - 19/03/2022

Horario: Estudios a distancia, Módulo/Asignatura sin horario

### Módulo/Asignatura 9. Deep Learning

Número de créditos: 4,00 ECTS

Modalidad de impartición: A distancia

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

Fechas de inicio-fin: 07/05/2022 - 10/06/2022

Horario: Estudios a distancia, Módulo/Asignatura sin horario

### Módulo/Asignatura 10. Inteligencia Colectiva y Sistemas de Recomendación

Número de créditos: 3,00 ECTS

Modalidad de impartición: A distancia

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: 15/01/2022 - 29/01/2022

Horario: Estudios a distancia, Módulo/Asignatura sin horario

### Módulo/Asignatura 11. Modelos para Datos Temporales y Espaciales

Número de créditos: 3,00 ECTS

Modalidad de impartición: A distancia

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: 15/04/2022 - 12/05/2022

Horario: Estudios a distancia, Módulo/Asignatura sin horario

### Módulo/Asignatura 12. Data Science & Business Intelligence

Número de créditos: 3,00 ECTS

Modalidad de impartición: A distancia

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: 25/03/2022 - 14/04/2022

Horario: Estudios a distancia, Módulo/Asignatura sin horario

### Módulo/Asignatura 13. Procesamiento del Lenguaje Natural para la Ciencia del Dato

Número de créditos: 4,00 ECTS

Modalidad de impartición: A distancia

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, Automatas 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: Automatas. 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: 03/02/2022 - 17/03/2022

Horario: Estudios a distancia, Módulo/Asignatura sin horario

### Módulo/Asignatura 14. Visualización de Datos

Número de créditos: 2,00 ECTS

Modalidad de impartición: A distancia

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

Fechas de inicio-fin: 18/02/2022 - 03/03/2022

Horario: Estudios a distancia, Módulo/Asignatura sin horario

### Módulo/Asignatura 15. Visualización de la Información

Número de créditos: 2,00 ECTS

Modalidad de impartición: A distancia

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: 20/01/2022 - 10/02/2022

Horario: Estudios a distancia, Módulo/Asignatura sin horario

### Módulo/Asignatura 16. Trabajo Fin de Máster

Número de créditos: 6,00 ECTS

Modalidad de impartición: A distancia

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: 01/03/2022 - 24/06/2022

Horario: Estudios a distancia, Módulo/Asignatura sin horario