

DATA SCIENCE Y BIG DATA (I EDICIÓN)

Datos básicos del Curso	Curso Académico	2015 - 2016
	Nombre del Curso	Data Science y Big Data (I Edición)
	Tipo de Curso	Diploma de Especialización
	Número de créditos	33,00 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, 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).
	Requisitos académicos para la obtención del Título o Diploma	Superar satisfactoriamente los 33 ECTS
	Criterios de selección de alumnos	
Preinscripción	Fecha de inicio	01/12/2015
	Fecha de fin	20/01/2016
Datos de Matriculación	Fecha de inicio	04/01/2016
	Fecha de fin	20/01/2016
	Precio (euros)	653,00 (tasas incluidas)
	Pago fraccionado	No
Impartición	Fecha de inicio	28/01/2016
	Fecha de fin	15/07/2016
	Modalidad	Presencial
	Idioma impartición	Español
	Lugar de impartición	

Laboratorio Módulo H1, Facultad de Matemáticas

Prácticas en
empresa/institución No

Información

Teléfono 954557879

Web <http://www.glc.us.es/masterdata>

Facebook

Twitter

Email jborrego@us.es

DATA SCIENCE Y BIG DATA (I 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 multidisciplinarios, 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 creación y gestión de conocimiento.

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

Procedimientos de Evaluación

Asistencia, Trabajos

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. Institución no universitaria - 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

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

D. Gonzalo Antonio Aranda Corral. Universidad de Huelva- Tecnologías de la Información
D^a. 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. José Manuel Camacho Sosa. - Consultor Informático

D^a. M^a Dolores Cubiles de la Vega. Universidad de Sevilla - Estadística e Investigación Operativa

D. Jaime de Miguel Rodríguez. Sushant School of Art & Architecture- Architecture

D. Juan Galán Páez. - Consultor de Datrik Intelligence S.L.

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

D. Miguel Ángel Gutiérrez Naranjo. Universidad de Sevilla - Ciencias de la Computación e Inteligencia Artificial

D^a. María José Hidalgo Doblado. Universidad de Sevilla - Ciencias de la Computación e Inteligencia Artificial

D^a. María Dolores Jiménez Gamero. Universidad de Sevilla - Estadística e Investigación Operativa

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

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

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

D. Gabriel Muñoz Ríos. - Consultor

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

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

D. José Francisco Quesada Moreno. - Centro de Informática Científica de Andalucía

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. - Consultor de Datrik Intelligence S.L.

Módulos/Asignaturas del Curso

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

Número de créditos: 3,00 ECTS

Modalidad de impartición: Presencial

Contenido: Introducción a la Ciencia del Dato

Experimentación e Investigación: Reproducibilidad y documentación
Proyectos básicos de aplicación de técnicas de Ciencia del Dato
Big data

Etapas de un proyecto

Exploración de datos
Técnicas de educación
Modelización: tipos de datos
Computación
Aplicación
Gestión de versiones
Elaboración y redacción de informes finales

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: 24/04/2016 - 25/05/2016

Horario: Miércoles
En horario de tarde

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

Número de créditos: 4,00 ECTS

Modalidad de impartición: Presencial

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 esestructurados
Optención de datos
Almacenamiento y consulta
Biblioteca estándar Python:
Otras bibliotecas:

Fechas de inicio-fin: 28/01/2016 - 18/03/2016

Horario: Jueves, Viernes
En horario de tarde

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

Número de créditos: 4,00 ECTS

Modalidad de impartición: Presencial

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: 28/01/2016 - 18/03/2016

Horario: Jueves, Viernes
En horario de tarde

Módulo/Asignatura 4. Aprendizaje Estadístico y Modelización

Número de créditos: 6,00 ECTS

Modalidad de impartición: Presencial

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: 31/03/2016 - 03/06/2016

Horario: Jueves, Viernes
En horario de tarde

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

Número de créditos: 4,00 ECTS

Modalidad de impartición: Presencial

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: Column Stores, BD Documentos, Clave-Valor, BD en grafo, etc.
- * Procesamiento en paralelo y escalabilidad (multiprocesamiento a nivel de datos VS a nivel funcional)
- * Casos de uso y diseño de soluciones
- * Entornos integrados. Ejemplo: RapidMiner

Fechas de inicio-fin: 31/03/2016 - 03/06/2016

Horario: Jueves, Viernes
En horario de tarde

Módulo/Asignatura 6. Machine Learning I

Número de créditos: 5,00 ECTS

Modalidad de impartición: Presencial

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: 08/06/2016 - 15/07/2016

Horario: Miércoles, Jueves, Viernes
En horario de tarde

Módulo/Asignatura 7. Inteligencia Colectiva, Big Data y Sistemas de Recomendación

Número de créditos: 3,00 ECTS

Modalidad de impartición: Presencial

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: 19/05/2016 - 17/06/2016

Horario: Jueves, Viernes
En horario de tarde

Módulo/Asignatura 8. Procesamiento del Lenguaje Natural para la Ciencia del Dato I

Número de créditos: 2,00 ECTS

Modalidad de impartición: Presencial

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.

Fechas de inicio-fin: 23/06/2016 - 08/07/2016

Horario: Jueves, Viernes
En horario de tarde

Módulo/Asignatura 9. Visualización de la Información

Número de créditos: 2,00 ECTS

Modalidad de impartición: Presencial

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: 08/06/2016 - 15/07/2016

Horario: Miércoles, Jueves, Viernes
En horario de tarde