



**4º EDICIÓN**

**ntic**  
master  
revolucionamos la comunicación

**Título Propio Universidad Complutense de Madrid**

Facultad de Estudios Estadísticos  
Universidad Complutense de Madrid

# **BIG DATA & BUSINESS ANALYTICS**



# Índice Programa

## Big Data

### MÓDULOS

**Módulo I:** Introducción al Big Data

**Módulo II:** Bases de Datos SQL

**Módulo III:** Business Intelligence

**Módulo IV:** Programación en Python

**Módulo V:** Bases de Datos NoSQL

**Módulo VI:** Tecnologías del Big Data

**Módulo VII:** Hadoop y Spark

**Módulo VIII:** Deep Learning

## MÓDULOS

**Módulo IX:** Programación en R

**Módulo X:** Fundamentos de Estadística

**Módulo XI:** Minería de Datos y  
Modelización Predictiva

**Módulo XII:** Machine Learning

**Módulo XIII:** Aplicaciones del Big Data  
en la Empresa

**Módulo XIV:** Data Science Aplicada a la empresa

**Módulo XV:** Text Mining y Redes Sociales

**Módulo XVI:** Open Data

**Módulo XVI:** Emprendimiento en Empresas  
de Big Data

**Módulo XVII:** Scala





# La importancia del **BIG DATA**



## **Empresas y organismos están comenzando a adaptarse a la nueva era**

Las empresas y organismos oficiales ya se están adaptando a los nuevos tiempos en los que la información masiva se procesa y analiza, convirtiéndose en un importante activo para la gestión empresarial en todas sus áreas de decisión. Lo que hace unos años comenzó siendo una ventaja competitiva de unos pocos, ahora está muy presente y, en breve, será imprescindible para no quedarse atrás: el dato es el petróleo del siglo XXI.

## **Macrodatos**

Los macrodatos son una valiosa herramienta en la creación de informes estadísticos, la identificación de nuevas oportunidades de negocio, modelos de predicción sobre los resultados publicitarios de una campaña futura o la evaluación de datos masivos para avanzar en investigaciones médicas que ayuden a erradicar enfermedades.

## **Aplicaciones prácticas**

La información a gran escala no es un factor que afecte únicamente al campo matemático o estadístico, pues sus aplicaciones prácticas abarcan todo tipo de entornos reales en los que, se torna fundamental para afrontar situaciones muy diversas a nivel empresarial, gubernamental, científico o social.





**Duración:**  
1 año académico

**Modalidades:**  
Presencial y Online

**Créditos ECTS:**  
60

**Modalidad Clases Presenciales:**  
Facultad de Estudios Estadísticos UCM  
**Viernes**  
16:00 – 21:00h  
**Sábado**  
9:00 – 14:00h

**Modalidad Online**  
100% online

## ¿Por qué estudiar en la UCM?

La Universidad Complutense de Madrid tiene más de 80 títulos de Grado y doble Grado, más de 140 másteres, además de estudios de formación permanente. La UCM tiene más de 500 años de historia y reconocimiento social. La Universidad Complutense de Madrid es la universidad española de referencia en 5 continentes.

El prestigio de la universidad está avalado por 7 Premios Nobel, 20 Príncipes de Asturias, 7 Premios Cervantes, Premios Nacionales de Investigación y a la Excelencia. La Universidad Complutense de Madrid tiene estudiantes de más de 90 países y convenios con universidades de los 5 continentes.



## ¿Por qué estudiar un título propio de la UCM?

Si hay algo que afianza los conceptos teóricos de un programa educativo es la práctica. Nuestros módulos formativos combinan una base teórica con ejercicios prácticos basados en situaciones reales de las empresas.

La preparación del Trabajo Final de Máster (TFM) garantiza la puesta en práctica de todos los conceptos adquiridos a lo largo del curso, capacitando definitivamente al alumno para asumir responsabilidades dentro de un entorno laboral real.



## Convenios con empresas

La Universidad Complutense de Madrid y el Máster en Big Data y Business Analytics mantienen convenios con prestigiosas empresas del sector, lo que permite a los alumnos acceder a prácticas con las que aplicar los conocimientos adquiridos a lo largo del máster.

El claustro de profesores de este Máster presenta la gran ventaja de reunir destacados profesores universitarios de prestigio con grandes profesionales en activo en compañías de referencia en el ámbito empresarial.



Programa máster

# Big Data y Business Analytics



## Módulo I: Introducción al Big Data

Una primera aproximación a los fundamentos de la inteligencia de datos y su aplicación en distintos entornos empresariales. Incluye las tendencias actuales de los macrodatos y una iniciación a la línea de comandos. A lo largo del primer módulo del máster se realizará una aproximación inicial a los fundamentos de la inteligencia de datos y su aplicación en la empresa y las organizaciones.

Conceptos y preguntas básicas en relación a los macrodatos como las diversas técnicas que almacenan y procesan los mismos, serán comentados en clase. También las tendencias actuales del Big Data y ejemplos para entender su importancia en múltiples ámbitos del mundo real tan diferentes como la medicina, la publicidad, los avances científicos o la gestión gubernamental. Por supuesto, la utilización de los datos masivos en el entorno empresarial contará con un especial protagonismo, siendo un foco de atención prioritario a lo largo de todo el máster. Hecha la presentación sobre la importancia del Big Data, se reflexionará igualmente con los alumnos sobre los retos y oportunidades que plantea en el presente y futuro cercano, realizando una breve introducción a las tecnologías de los datos masivos.

El primer capítulo del curso, Finaliza con una introducción a la línea de comandos que proveerá del interfaz necesario para facilitar la interacción entre equipo informático y usuario.

## Módulo II: Bases de Datos SQL

## Índice de Contenidos

Diseño y modelización de base de datos y lenguaje de consulta estructurada, más conocido como SQL (Structured Query Language). Estudio del modelo entidad-relación, modelo relacional, e implementación relacional con SQL.

Los estudiantes tendrán como objetivo general en este módulo adquirir los conceptos fundamentales de las bases de datos y sus técnicas básicas de diseño, gestión y explotación, haciendo hincapié en el modelo de bases de datos relacionales.

Se realizarán prácticas para asentar correctamente estos conocimientos, pues se trata del modelo de mayor implantación en la actualidad, y sus conceptos y técnicas están presentes en todos los demás.

Podemos afirmar que las bases de datos SQL son el punto de partida para el manejo de volúmenes de datos, ya sean pequeños o grandes, y por tanto una parte crucial en la iniciación del proceso que lleva a sacar el máximo rendimiento a la inteligencia de datos para ponerla al servicio de un objetivo concreto.

Además de las técnicas conceptuales, en este módulo se estudia el lenguaje SQL para la definición, consulta y manipulación del dato. A continuación, se realizará un resumen de los contenidos que se van a tratar a lo largo del módulo en mayor detalle.



UNIVERSIDAD  
COMPLUTENSE  
MADRID

Introducción a las bases de datos, sus sistemas de gestión y ventajas. Se profundizará en qué es una base de datos, su evolución y tipos.

El modelo entidad/relación en el diseño de las bases de datos, incluyendo el modelado conceptual de datos, las entidades, especialización y generalización, relaciones, cardinalidad y atributos.

Conceptos fundamentales del modelo relacional como relaciones, claves y restricciones de integridad, así como normalización y transformaciones del modelo entidad/relación al relacional.

Implementación relacional con SQL: definición de datos (DDL), manipulación de datos (DML), consultas simples, subconsultas, Join Exist y Not Exist, Having y Group By, Union, Intersect y Except, Insert, Update y Delete.

Conceptos fundamentales y arquitectura de bases de datos, SQL Server Management Studio, lenguaje Transact-SQL, transacciones, seguridad, vistas, triggers, DDL, DML, Stored Procedures y funciones.





## Módulo III: Business Intelligence con Tableau

Este módulo plantea diferentes objetivos al alumno, con los que espera mejorar sus capacidades analíticas, así como sus habilidades para desenvolverse en un entorno empresarial dentro del ámbito del “Business Intelligence”.

El módulo introduce al alumno al concepto de Business Intelligence, diferenciando éste, del concepto de MachineLearning o de DataScience, su relación con los nuevos paradigmas de BigData.

Para pasar a un enfoque completamente práctico en el que el alumno aprenderá a utilizar Tableau. Se acompañará al alumno en el proceso de descubrimiento de claves (insights) aplicado sobre un conjunto de datos abiertos. Tableau es líder en el cuadrante de visionarios de Gartner en las plataformas de BI en los últimos cinco años de forma consecutiva (recientemente fue adquirida por Salesforce). Con esta introducción el alumno podrá realizar análisis básicos usando esta solución que complementarán al aprendizaje de otros lenguajes y técnicas analíticas de este Máster.

Como complemento de esta vertiente técnica, el alumno aprenderá otros conceptos/habilidades más orientados a cómo desenvolverse en un entorno empresarial orientado a la analítica avanzada. Uno de estos conceptos se centra en cómo han de usarse de forma efectiva y eficiente diferentes tipos de gráficos. Y de cómo preparar y comunicar de forma eficiente los resultados de un análisis de datos a una audiencia no-técnica, de negocio. Para los analistas de negocio estos dos elementos se están considerando como esenciales en las organizaciones.



## Módulo IV: Introducción y Fundamentos de Programación en Python

Características, tipos de datos, estructuras de control de flujo, funciones, parámetros, manipulación de cadenas, estructuras de datos...

El objetivo general de este módulo es adquirir los conceptos fundamentales para sentar las bases que permiten dominar el lenguaje de programación Python, administrado por la Python Software Foundation bajo código abierto.

Esto significa que permite modificaciones de la fuente del programa sin restricciones, lo que le dota de un gran potencial gracias a los múltiples equipos de desarrolladores que trabajan en la mejora del mismo en el mundo.

Esa libertad y capacidad de evolución del propio lenguaje de programación Python, es probablemente uno de los factores por los que está teniendo una implantación fortísima en el ámbito del análisis de datos, la ciencia de los datos y el Big Data. Su conocimiento es imprescindible para cualquiera que trabaje en estas áreas y otras relacionadas. Todas estas razones, le convierten en uno de los imprescindibles de nuestro programa del Máster en Big Data y Business Analytics.

Otro atractivo con el que cuenta este módulo es su enfoque práctico, Nuestro programa no solo incluye la teoría necesaria también la visión práctica para experimentar sus técnicas básicas en una amplia gama de aplicaciones.



UNIVERSIDAD  
COMPLUTENSE  
MADRID

### Índice de Contenidos

Dentro de esa base teórica imprescindible para la posterior aproximación a sus múltiples funciones, se verán contenidos como:

Introducción a la programación con Python y conceptos básicos como:  
variables, instrucciones generales, tipos de datos y operaciones. Estos conocimientos impartidos nos permitirá a continuación abordar otros aspectos como:

Entrada y salida.

Estructuras de control:  
selección, iterativas.  
Funciones. Recursividad. Orden superior en Python.  
Expresiones lambda, map y reduce.

Estructuras de datos fundamentales:  
colecciones, listas, tuplas, conjuntos, diccionarios.  
Programación orientada a objetos.

Librerías para el procesamiento numérico y estadístico y para el análisis de datos.



## Módulo V: Bases de Datos NoSQL

Se aprenderá a utilizar y modelar los sistemas de gestión de bases de datos NoSQL y sus principales operaciones. Introducción a MongoDB, operaciones CRUD, dominar el Find o proyectar los campos en resultados de búsqueda.

Cuando hablamos de bases de datos NoSQL, nos estamos refiriendo a aquellas cuya característica más destacable es que no requieren de estructuras fijas como tablas, a diferencia de las bases de datos tradicionales que requieren de SQL como lenguaje principal de consultas. Sin embargo, esto no significa que no soporten dicho lenguaje SQL.

### Introducción a las bases de datos NoSQL

- Bases de datos relacionales vs NoSQL (ACID vs. BASE), donde los alumnos podrán ver las diferencias entre estos dos tipos de bases de datos.
- Modelo de datos (entidad relación vs agregación)
- Diferentes tipos de bases de datos NoSQL (key-value, XML, grafos, documentos, columnas). Se abordarán las características más destacadas de cada una de ellas
- Cuando utilizar NoSQL (teorema CAP)

### Mongo vs. Cassandra

- Modelo de datos
- No Joins
- Array y documentos embebidos
- Desnormalización

### Operaciones CRUD

- Create: insert, insertOne, insertMany.
- Update: update, updateOne, updateMany, findAndModify

### Introducción a MongoDB

- Como instalar MongoDB en Windows/Mac/Linux donde se verá todo el proceso paso a paso.
- Cómo conectarse a la shell (vía terminal/RoboMongo).
- Crear/borrar base de datos/colecciones.
- Copias de seguridad y restauración de este sistema de base de datos NoSQL orientado a documentos.

### Dominar el Find

- Aplicar filtros avanzados.
- Ordenar los resultados de búsqueda.
- Pagar los resultados de búsqueda.
- Limitar los resultados de búsqueda.

### Proyección, Indexes & Aggregation Cursores

- Proyectar los campos en los resultados de búsqueda.
- Ensure index y full text search.
- Aggregation (sum, avg...)
- Cursores.



## Módulo VI: Tecnologías del Big Data

Internet de las cosas como servicio, conectividad con fuentes de datos heterogéneas a través de brókers de mensajes y hubs con dispositivos, Hadoop, Spark y diversas técnicas de visualización y análisis de información por parte de los usuarios finales.

### **Internet of things as a Service**

Aproximación a la aportación del Internet de las cosas al mundo del macrodato. El alumno aprenderá a descubrir y valorar oportunidades presentes en el día a día del ciudadano y de la empresa.

### **Conectividad con fuentes de datos**

heterogéneas a través de bróker de mensajes y hubs con dispositivos

- ETL as a Service.
- Gobierno de los datos as a Service.
- Stream Analytics, CEP análisis en memoria en tiempo real en los eventos complejos as a Service.

### **Spark**

- Introducción a Spark, sistema de computación que tiene en la velocidad su característica destacada.
- Sistemas de caché y persistencia.
- Cluster Spark.
- Desarrollo de aplicaciones con Spark.
- RDD y transformaciones.
- Spark Streaming.

### **Hadoop**

- Introducción a Hadoop, entorno de trabajo que soporta diversas aplicaciones distribuidas bajo una licencia libre. En el ámbito de la inteligencia de datos ostenta un gran peso dentro los principales programas existentes.
- Instalación y configuración paso a paso.
- Almacenamiento HDFS infinito en Apache Hadoop.
- BBDD MPP offering as a Service.
- Análisis estadístico y aprendizaje automático PaaS.
- Google BigTable y Hbase.
- Transformación de datos con Apache Hadoop.
- Explotación de datos con Apache Hadoop para sacar el máximo rendimiento a este entorno de trabajo.

### **La visualización y análisis de información por parte de los usuarios finales.**

- Modelos de inteligencia cognitiva.
- PowerBI, cuadros de mando en tiempo real.



## Módulo VII: Hadoop y Spark

En este módulo repasaremos las tecnologías Big Data y su motivación en el contexto actual de la era digital y las necesidades de las empresas, fundamentalmente la personalización y la orientación a cliente o customer-centricity. A continuación nos adentraremos en dos tecnologías del ecosistema Big Data actual como son HDFS (Hadoop Distributed File System) y Apache Spark.

El alumno podrá adquirir una visión panorámica de HDFS, su arquitectura y su utilización a través de línea de comandos. Es el sistema de almacenamiento fundamental en el mundo Big Data en la actualidad, por lo que es imprescindible que el alumno conozca y experimente su funcionamiento.

Pasaremos después al estudio del tema central del curso, Apache Spark, sin duda la tecnología más demandada en la actualidad para procesamiento de grandes volúmenes de datos. Describiremos su filosofía y enfoque para ejecutarse sobre un cluster de ordenadores, e iremos desgranando cada uno de los módulos que lo componen, con especial énfasis en los módulos de Spark SQL y Spark MLlib, dos de las piezas clave en el día a día de un Data Scientist en la actualidad. Se usará la infraestructura de Google Cloud para que cada alumno pueda desplegar de forma sencilla un cluster de Spark bajo demanda que utilizaremos durante las clases.



## Módulo VIII: Deep Learning

El contenido de la sección de deep learning estará estructurado en cuatro bloques. La duración de los mismos será de unas dos o tres horas, pudiendo ajustarse a la audiencia y dependiendo del aprendizaje de los alumnos. Los conceptos más básicos son los de los dos primeros bloques, mientras que los dos últimos estarán más orientados a la profundización. Durante las clases se impartirá una componente práctica en cada bloque para familiarizar a los alumnos con la aplicación directa.

La estructura será la siguiente:

### **Bloque 1: Redes Neuronales**

En este bloque se pretende introducir el concepto de red, explicando su funcionamiento y permitiendo que el alumno comprenda su funcionamiento a nivel teórico y práctico. Los contenidos de este bloque son los siguientes:

- Introducción: “from Representation Learning”.
- Forward and backward propagation.
- Descenso gradiente. Batches y online training.
- PRÁCTICA: Implementación de una red neuronal desde numpy.
- Tensores y Frameworks para deep learning: Keras (Tf). Pytorch.
- Funciones de activación, optimizadores y funciones de coste. Hiperparámetros en una red neuronal.

### **Bloque 2: Redes Convolucionales. Imágenes.**

En este segundo bloque se presentará una de las estructura más comunes en el mundo del deep learning: las redes convolucionales. Se hará hincapié en su capacidad de generar features para el modelo desde estructuras de datos no tabulados, tales como imágenes.

- Extracción de características. Convoluciones.
- Pooling y padding. Efectividad de la activación Relu.
- Conexión con la capa densa (clasificador).
- Arquitecturas de red. Imagenet.
- PRÁCTICA: Se construirá un clasificador de imágenes sencillo usando keras.
- Introducción a Transfer Learning.

### **Bloque 3: Redes Recurrentes**

Durante este bloque se presentará el enfoque de secuencias desde las redes neuronales, introduciendo ciertos mecanismos de memoria. Se presentará también la idea de embedding y se introducirá el tratamiento de texto.

- Secuencias temporales. Timesteps.
- RNN vainilla.
- Práctica: Series temporales.
- Vanishing gradient. LSTM.
- Redes recurrentes sobre secuencias de palabras.
- PRÁCTICA: Generación de texto a nivel de carácter.

### **Bloque 4: Embeddings, Autoencoders y Redes Generativas**

Durante este último bloque se profundizará un poco más en los resultados intermedios de las redes, introduciendo el concepto de embedding y las posibilidades al tratar el espacio de features como vectorial. Se introducirán los autoencoders y se presentará su variante variacional como red generativa.

- Reducción de dimensiones.
- Embeddings
- Word embeddings: word2vec, glove.
- Espacio Latente. Detección de anomalías
- Variational autoencoders
- PRÁCTICA: entrenamiento de un VAE para generación
- GAN

## Módulo IX: Introducción y Fundamentos de Programación en R

Fundamentos, estructuras de control y manejo de datos del lenguaje de programación R, muy útil en entornos como la minería de datos, estadística o matemáticas.

El lenguaje de programación en R es ampliamente utilizado con fines estadísticos, data mining, matemática financiera o incluso en bioinformática e investigación biomédica. Ello es debido a la posibilidad que ofrece de trabajar con diversos paquetes de funcionalidades gráficas y de cálculo.

Los estudiantes aprenderán a modelizar, construir y diseñar bases de datos multidimensionales, de tal manera que se agilicen sus consultas y puedan ser explotadas posteriormente según el objetivo del estudio.

También comprenderán la necesidad de estas técnicas, sus objetivos y aplicaciones, en función del tipo de información de que se dispongan en cada caso.

Gestión y modelización de bases de datos: creación, depuración y diseño de consultas. El alumno estudiará también el acceso a bases de datos mediante paquetes estadísticos de este lenguaje de programación, para lo que se utiliza extensiones que permiten añadir funcionalidades a la configuración básica del mismo.

### Contenido

- **Introducción al entorno R.** Se tratarán los aspectos generales y características más importantes a tener en cuenta de este lenguaje de programación.

- **Modo consola y modo script en R.**

- **Objetos en R.**

- **Estructuras de control de flujo de ejecución de la programación en R.**

- **Funciones en R.** Donde se realizará un repaso a las instrucciones características de este lenguaje de programación.

- **Gráficos en R.** Se profundizará en su versatilidad, los dispositivos gráficos necesarios para llevarlos a cabo, las ventanas gráficas, diagramas de barras, gráficos de contorno, gráficos 3D, rutinas de dibujo generales, histogramas, pares de gráficos de dispersión por variables y otras funcionalidades adicionales.

- **Paquetes de R.** Se estudiarán sus posibilidades: Manipulación de los datos, la carga de estos, modelización, visualización y presentación de resultados.



## Módulo X: Fundamentos de Estadística

Consolidación de los conocimientos de estadística necesarios para adquirir una base de conocimiento que ayudará a seguir el resto de bloques del máster. Entre los conceptos a tratar, la estadística descriptiva y la probabilidad e inferencia tendrán un destacado espacio.

Gracias a la estadística se pueden reunir, organizar y analizar diversos datos muy útiles para plantear una base sobre la que tomar decisiones en múltiples ámbitos. Un modelo muy práctico para la resolución de diversos tipos de problemática y la realización de modelos predictivos.

En este módulo del Máster en Big Data y Business Analytics de la Universidad Complutense de Madrid, los alumnos serán formados para que, a su finalización posean una base de conocimiento y práctica que permita avanzar hacia la finalización del programa de forma exitosa.

### **Estadística descriptiva:**

- Descripción de variables estadísticas univariantes. Centradas en una característica en particular del objeto de estudio, pueden ser numéricas o de otra índole (como por ejemplo sexo, nivel de estudios o sector profesional).
- Se profundizará en diversos tipos de medidas estadísticas de centralización, dispersión, asimetría y curtosis.
- Descripción de variables bidimensionales. A diferencia de las univariantes, tienen en cuenta dos caracteres del mismo sujeto de estudio y las posibles relaciones entre dos objetos distintos.
- Análisis de la vinculación.
- Medidas de asociación.
- Regresión.

### **Inferencia:**

Cuando en el campo estadístico hablamos de inferencia, nos estamos refiriendo a las diversas técnicas y metodologías por las que, en base a una información con la que previamente contamos, se realizan modelos de predicción sobre el comportamiento de un conjunto determinado de población. Visión en la que se contempla además un posible margen de error medible.

Un ejemplo muy cotidiano al respecto son las encuestas electorales de intención de voto. Partiendo de una base de encuestados que busca ser lo más heterogénea posible, se intenta determinar cuál será el reparto de votos entre los distintos partidos políticos que se presentan a unas elecciones.

Algunos de los puntos que se tratarán en este módulo en relación a la inferencia estadística serán:

- Variables aleatorias.
- Modelos de distribución de probabilidad.
- Estimulación puntual de parámetros.
- Estimulación por intervalos de confianza.
- Contrastes paramétricos.
- Contrastes no paramétricos.





## Módulo XI: Minería de Datos y Modelización Predictiva con R

A lo largo de este bloque, los alumnos adquirirán los conceptos necesarios para el desarrollo de la modelización predictiva. Para ello, detectarán patrones basados en grandes volúmenes de datos a través de diversas técnicas de data mining.

Los alumnos del Máster en Big Data y Business Analytics de la Universidad Complutense de Madrid, aprenderán a través de este módulo las diversas técnicas de minería y modelos predictivos básicos, y aplicaciones en credit scoring. Los resultados de su aplicación serán de gran utilidad en múltiples tareas posteriores, de las cuales, se ofrece más información con ejemplos a continuación.

Gracias a los conocimientos adquiridos, los estudiantes podrán descubrir patrones en conjuntos de macrodatos, transformando estos en estructuras que sean comprensibles para su posterior análisis y uso en alineación a los objetivos empresariales o de otra índole que procedan.

Entre otros, abarcarán los siguientes puntos:

- Integridad y depuración de datos. Este es un punto fundamental para muchos propósitos en los que se aplica la inteligencia de datos. Así, evitar errores o información incorrecta entre aquella con la que contamos, reportará, entre otros, un retorno en factores como mayor productividad o rentabilidad dentro del entorno empresarial.
- Regresión lineal y logística.
- Técnicas de reducción de la dimensionalidad.
- Análisis y predicción con series temporales.
- Clasificación no supervisada.
- Análisis cluster.
- Construcción de scorecard, modelo de gestión y planificación ampliamente utilizado en el mundo empresarial, sobre el cual profundizaremos a lo largo de este módulo del máster.



## Módulo XII: Machine Learning con R y Python

Técnicas y aplicaciones de aprendizaje y modelización predictiva avanzada. Posteriormente, se profundizará en redes Deep Learning con R y Python, así como en modelos predictivos basados en árboles de decisión, Random forest, Gradient Boosting y Xgboost con R.

Los alumnos aprenderán diversas herramientas y aplicaciones de Machine Learning y modelización predictiva avanzada.

Estas técnicas de aprendizaje automático permiten a las máquinas ser capaces de asimilar una serie de comportamientos generalizados para realizar diversas acciones que toman, como ejemplo de referencia, las pautas indicadas previamente.

Es importante ser conscientes, y así se les hace saber a nuestros estudiantes, que el Machine Learning está presente en buena parte de elementos con los que interactuamos diariamente en nuestra vida cotidiana.

Teléfonos móviles o sistemas de navegación, por poner únicamente dos ejemplos, cuentan con este sistema de aprendizaje automático. Igualmente, aplicaciones de detección de fraudes, diagnósticos médicos o el propio buscador de Internet --que te devuelve resultados a una consulta-- tienen presente el mismo.



UNIVERSIDAD  
**COMPLUTENSE**  
MADRID



Debido a la importancia de este módulo dentro de los que componen el temario del máster, se repasarán a lo largo del mismo los siguientes puntos:

- Introducción.
- Redes neuronales y Deep Learning.
- Árboles de decisión.
- Random forest.
- Gradient Boosting.
- Support Vector Machines.
- Algoritmo KNN.

Dentro del aprendizaje automático, las tecnologías usadas para llevar a cabo proyectos de Machine Learning tienen igualmente y como es lógico un gran protagonismo.

Es por ello que el programa en este punto no se detendrá únicamente a dar a conocer a los alumnos las características más comunes del aprendizaje automático, sino que pondrá también en su mano las herramientas necesarias para poder desarrollar su propio proyecto al respecto.

## Módulo XIII: Aplicaciones del Big Data en la Empresa

Este módulo tiene como objetivo que los alumnos asimilen los distintos sistemas de soporte a la toma de decisiones en un entorno corporativo. Se abarcarán las áreas de gestión económica y financiera, operaciones, logística, marketing y ventas.

Una de las características destacadas es estar muy orientado a la aplicación de la inteligencia de datos en entornos empresariales. Por lo tanto, este módulo de Aplicaciones del Big Data en la empresa toma un gran protagonismo para la visión eminentemente práctica que se traslada al alumno.

El objetivo principal de este capítulo del máster es enseñar a los alumnos las aplicaciones del Big Data dentro de un entorno empresarial, abarcando diferentes aspectos y funcionalidades de los macrodatos asociados a departamentos y características como:

- Sistemas de soporte a la decisión.
- Marketing y ventas.
- Gestión económica financiera.
- Operaciones y logística.

Y hay que tener en cuenta que estos son solo una pequeña muestra de todos los departamentos y factores a tener en cuenta, como veremos a continuación.

Los macrodatos son decisivos en los procesos de múltiples áreas de gestión dentro de una organización empresarial. Gracias a ellos, corporaciones líderes, y por supuesto otras más modestas, han logrado mejorar sus procesos de producción, incrementar sus ingresos, reducir gastos, implementar campañas publicitarias, mejorar la seguridad de sus instalaciones, facilitar un mejor servicio de atención al cliente y realizar análisis predictivos más acertados. En definitiva, tomar mejores decisiones.

Lo realmente interesante es saber que, bien gestionada, es muy grande la capacidad que tienen los macrodatos para incidir de forma positiva en procesos y departamentos tan diversos dentro de un entorno corporativo o industrial. De ahí su importancia y su verdadero factor diferencial respecto a otras técnicas y metodologías.



## Módulo XIV: Data Science Aplicada a la Empresa

Visión integral de las empresas orientadas al dato, creación de equipos de científicos de datos y estructuración de un proyecto Data Science. Este módulo aborda igualmente aspectos de la comunicación personal con individuos, la pública ante grupos, y la mediática para audiencias.

El objetivo principal es proporcionar a los alumnos una visión integral de las empresas orientadas al dato, formarles en la creación de equipos de Data Science y enseñarles a estructurar un proyecto de ciencia de datos.

Es importante que los estudiantes conozcan el amplio campo de acción de esta nueva ciencia y sus ventajas y aplicaciones prácticas en una estrategia digital a nivel global dentro de entornos empresariales. Los procesos, métodos científicos y sistemas estarán muy presentes a lo largo de este bloque del programa del máster.

El perfil del Data Scientist es crucial en todo el proceso que envuelve al macrodato. Es habitual que sea el científico de datos el profesional encargado de sacar el máximo provecho a los datos de los que dispone la empresa.

Contar con una excelente capacidad de análisis es muy útil en el perfil que nos ocupa, pero también es interesante que disponga de los conocimientos necesarios para, a través de las herramientas existentes a tal efecto, llegar a esas conclusiones analíticas.

El Data Science permite ir un paso más allá en el campo del análisis de datos. Tras lo aprendido con los fundamentos estadísticos, el data mining, el aprendizaje automático o los métodos de análisis predictivo, es lo que ha otorgado mayor protagonismo a este campo de conocimiento en los últimos años: ha sido la evolución natural derivada de la relación aplicada del dato al objetivo corporativo de la empresa.



## Módulo XV: Text Mining y Redes Sociales

Utilización de las principales APIs de acceso a datos en redes sociales, análisis de texto y su tratamiento adecuado para su uso efectivo. También se realizará el análisis de sentimiento en redes sociales y las distintas formas de representación de la información obtenida de éste.

### Text mining

En el bloque de minería de textos, que forma parte del módulo text mining y redes sociales, se pretende instruir a los alumnos en el procesamiento de textos como forma de análisis de información no estructurada o semiestructurada.

Se estudiarán conceptos propios de este campo que tanto ha avanzado en los últimos años. Así, algunos aspectos a tratar serán: la extracción de textos de distintos tipos de fuentes web; preprocesamiento (limpieza, transformación, obtención de raíces, etc.); exploración y procesamiento (agrupación, modelos temáticos, minería de opiniones y análisis de sentimiento).

En el desarrollo de este bloque se utilizará el lenguaje R, pudiéndose usar –en función de los intereses del alumnado– otros lenguajes adicionales como Python. Si así lo aconsejan las inquietudes de los estudiantes, se podrá acceder a dos perspectivas desde distintos lenguajes de programación para llevar a cabo un trabajo de text mining.

### Redes sociales y Big Data

En el segundo bloque del módulo, el de redes sociales en relación al Big Data, comenzaremos haciendo un análisis de las propiedades de una red social desde la perspectiva de los macrodatos, tratando aspectos diversos como la densidad, tamaño ó diámetro.

Se continuará con una clasificación, indicando claramente si estamos ante una red aleatoria o con estructura.

Tras esto, se pasará a medir la centralidad de los distintos agentes involucrados en la red.

Para realizar todos estos análisis se utilizará el software libre PAJEK, que permite tanto manejo de redes de gran tamaño, como de menor entidad.

La combinación de los dos bloques que componen el módulo permitirán a los alumnos una visión global en relación tanto a la parte más teórica del temario como a sus aplicaciones en entornos prácticos.



## Módulo XVI: Open Data

Tendencias en gobierno abierto, participación y transparencia se tratarán en este bloque. Del mismo modo, se abordará las fuentes datos abiertos incluidas, las no gubernamentales (APIs) y fuentes de algoritmos abiertos de última generación como pueda ser Kaggle. A continuación se verán ejemplos diversos de código reproducible.

Los datos abiertos son aquellos que están a disposición pública, sin restricciones de uso ni necesidad de permisos. Tampoco están sujetos a patentes de ningún tipo.

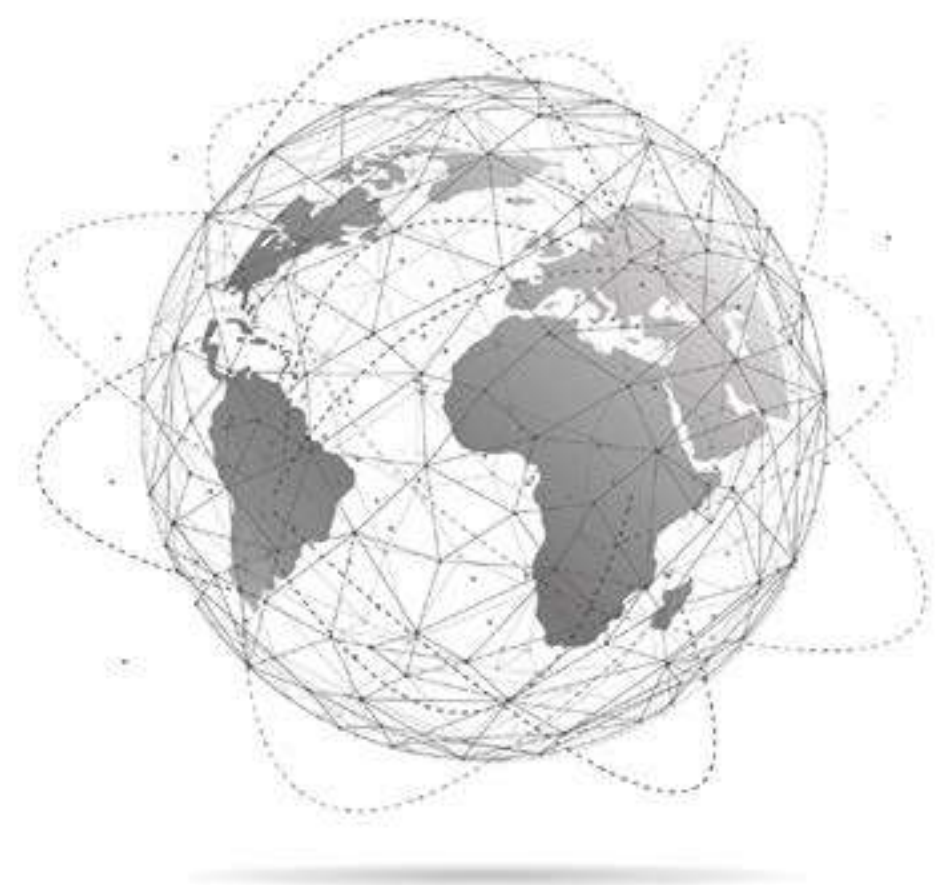
Se analizarán diversas fuentes de datos abiertos existentes en la actualidad, con un repaso a la tendencia creciente de permitir el acceso libre a datos tradicionalmente restringidos al ámbito corporativo y gubernamental.

Aprender no únicamente a acceder a esos datos, sin también saber cómo tratarlos para sacar de ellos conclusiones interesantes y factores de valor que estén alineados con nuestros intereses, será lo que realmente dote de utilidad a este tipo de datos más allá del componente meramente informativo.

En cualquier caso, la apertura de los datos al gran público no está exenta de polémicas y posiciones encontradas. Elementos a tener en cuenta son el posible uso de información personal sobre individuos específicos, ciertos datos sensibles que puedan ser considerados una amenaza a la seguridad si se dan a conocer. Son parte de un debate que genera interés y que transformará el mundo de los datos abiertos. Hay que monitorizar la evolución del dato abierto para saber cómo se gestiona y regula.

Entre los aspectos que se van a tratar en este módulo dedicado al dato abierto, encontramos:

- Open data gubernamental. Con las tendencias en gobierno abierto, participación y transparencia. Se expondrá el cambio que se está experimentando en muchas instituciones públicas que previamente restringían el acceso a sus datos y actualmente, bajo la premisa de actuar bajo una mayor transparencia, facilitan parte de la información de la cual disponen, tradicionalmente de uso interno, para su consulta pública.
- Otras fuentes de datos abiertos. Identificando las principales fuentes, incluidas las no gubernamentales (APIs).
- Fuentes de algoritmos abiertos. Los concursos, en especial Kaggle, son una fuente de acceso a algoritmos de última generación.
- Ejemplos de código reproducible.





## Módulo XVII: Emprendimiento en Empresas Big Data

Este bloque enseñará a los alumnos el diseño y proceso para el lanzamiento de una startup basada en servicios de Big Data, algo que resultará de gran utilidad especialmente a los estudiantes que tengan en mente iniciar su propio proyecto empresarial en el ámbito de la inteligencia de datos

## Módulo XVIII: Scala

Este lenguaje de programación orientado a objetos es muy similar a Java, incluyendo características de lenguaje funcional. Dentro del mundo del Big Data se habla de Scala ya que Spark, que es una de las plataformas que se emplean para procesar datos de tipo Big Data, está diseñado con Scala.

## Trabajo Fin de Master

Asimilados todos los conceptos previos, llega el momento de poner a prueba todos los conocimientos adquiridos en el máster. El alumno planteará una estrategia global de inteligencia de datos para una empresa, basándose en diferentes técnicas y softwares de apoyo de entre los existentes en el mercado.



Equipo directivo

# MÁSTER BIG DATA & BUSINESS ANALYTICS



## Javier Portela García-Miguel

Director. Profesor Titular UCM, doctor en Ciencias Matemáticas UCM, licenciado en Ciencias Matemáticas UCM. En el campo de la docencia, dirige diversos estudios en Data Mining y Business Intelligence en la UCM.



## José Carlos Soto Gómez

Co-Director y Profesor Asociado de la UCM. Socio Fundador de NTIC Master y Aplimovil. Amplia experiencia en proyectos nacionales e internacionales en IT y analítica en empresas como Banco de España, NEC, Telefónica, Vodafone, Orange, medios de comunicación...



## Equipo docente del Máster Big Data

Contamos con verdaderos profesionales del sector como profesores del Máster de Big Data & Business Analytics.

Estos altos cargos en activo forman a nuestros alumnos en contenido actualizado, de calidad y demandado por las instituciones actuales.

***“Aprende con los mejores profesionales del Big Data y Business Analytics”***

# Equipo Docente

# PROFESORES



**Pedro Pablo  
Malagón Amor**

*Sales Engineer Google Cloud*

Tras 17 años trabajando en Microsoft como Jefe de Proyecto, Ingeniero o Cloud Data Architect, nuestro profesor de Tecnologías del Big Data, trabaja en Google aspectos relacionados con datos masivos y su uso empresarial.



**Carlos  
Ortega Fernández**

*Senior Data Scientist en Teradata*

Carlos trabaja como Senior Data Scientist en ThinkbigAnalytics, compañía del grupo Teradata, donde desarrolla capacidades avanzadas basadas en datos, algoritmos y Machine Learning en todo tipo de industrias.



**Cristóbal  
Pareja Flores**

*Catedrático EU en la UCM*

Con 30 años como docente, Cristóbal es matemático especializado en Ciencias de la Computación. Además es decano de Posgrado e Investigación y doctor en Informática.



**Aída  
Calviño Martínez**

*Docente e investigadora UCM*

Aída es actualmente personal docente e investigadora de la UCM. Ganadora en 2014 del prestigioso Premio Abertis, participará como profesora en dos módulos y también en la supervisión del Trabajo Final de Máster.



**Conrado Miguel  
Manuel García**

*Catedrático EU en la UCM*

Conrado es Director de Departamento en la Facultad de Estudios Estadísticos y ha participado activamente en grupos de investigación. Es el encargado del módulo de Fundamentos de Estadística.



**Javier  
Monjas**

*Analytical Project Manager Sanitas*

Javier es un referente en su sector. Invitado de forma asidua como conferenciante en prestigiosas universidades y escuelas de negocios, gestiona en Sanitas la Dirección de Proyectos Analíticos.



# Equipo Docente

# PROFESORES



**Lorenzo**

**Escot Mangas**

*Profesor titular de la UCM*

Además de su experiencia como docente en la Complutense los últimos 20 años, Lorenzo es Codirector del Grupo de Investigación Análisis Económico de la Diversidad y Políticas de Igualdad de la UCM.



**Juana María**

**Alonso Revenga**

*Profesora titular de la UCM*

Profesora de la Universidad Complutense de Madrid desde 1985, es experta de SPSS, SAS y Estadística. Es una de las docentes del módulo de «Minería de datos y modelización predictiva con R».



**Daniel**

**Gómez González**

*Profesor titular en la UCM*

Coordinador del programa de doctorado en Data Science, Daniel es experto en SPSS, estadística y ciencias de la computación. Guía a los alumnos a través del módulo inicial de Introducción al Big Data.



**Santiago**

**Mota Herce**

*Corporate Advisor*

Santiago es consultor freelance en Business Intelligence, Machine Learning y estrategia. Ha asesorado a empresas como Bankia, Vodafone, Teradata o The Boston Consulting Group.



**María Isabel**

**Ríomos Callejo**

*Docente UCM*

Coordinadora del Campus Virtual (Facultad de Estudios Estadísticos). Delegada del Decano para funciones de asesoramiento sobre asuntos informáticos y nuevas tecnologías.



**Álvaro**

**Bravo Acosta**

*Ingeniero Técnico Informático en Sistemas*

Experto en Tecnologías Big Data, BI y Analítica. Desde octubre de 2008 ha trabajado en diferentes consultoras como Minsait, Sopra Steria o Everis, para clientes externos como ISBAN y BBVA.



# Equipo Docente

# PROFESORES



**José Luis Brita**

*Docente en la UCM*

Además de profesor en la UCM, cuenta con historial como investigador en proyectos de innovación educativa. Está a cargo del módulo de Introducción y Fundamentos de Programación en R.



**Pablo J. Villacorta**

*Científico de datos en Stratio*

Doctor en Ciencias de la Computación e IA, Ingeniero informático y licenciado en Estadística por la Univ. de Granada. Desarrollador Certificado en Spark 2.x por Databricks y autor de varios paquetes de R publicados en CRAN.



**Javier Castro Cantalejo**

*Docente en la UCM*

Más allá de la docencia, tiene experiencia como miembro de grupos de investigación. En el Máster de Big Data y Business Analytics de la UCM, es uno de los profesores del módulo de Text Mining y Redes Sociales.



**Ismael Yuste**

*Strategic Cloud Engineer en Google*

Trabaja como Strategic customer en EMEA. Experto en Big Data y GSuite. Bigquery, Dataproc, Dataflow, DataStudio, Pub/Sub, Datalab, DataPrep. Trabaja en Google Apps & Cloud como Customer Success Engineer.



**Fernando Velasco**

*Investigador de técnicas y algoritmos de inteligencia artificial*

Fernando posee un Máster en métodos matemáticos avanzados aplicados en físicas. Además, estudia las interacciones humanas mediante Deep Learning.



**Charles Flores Espinoza**

*Big Data Engineer en Stratio Big Data*

Experto Scala y Python. Charles está avalado por sus más de diez años de experiencia en el sector. En la actualidad, además estudia Data Scientist.



# Equipo Docente

# PROFESORES



**Javier Portela  
García-Miguel**

*Profesor Titular UCM*

Doctor en Ciencias matemáticas UCM y licenciado en Ciencias Matemáticas UCM. En el campo de la docencia, dirige diversos estudios en Data Mining y Business Intelligence en la UCM.



**Rosa  
Espinola**

*Doctora en la UCM*

Cuenta con sexenio como investigadora. Es miembro de proyectos de investigación financiados por el Ministerio de Educación y Ciencia para técnicas de Machine Learning.



**Luis  
Gascó Sánchez**

*Data Scientist*

Doctor por la Universidad Politécnica de Madrid con experiencia internacional en instituciones de I+D+i. Experto analizando datos de entornos urbanos aplicando técnicas de Machine Learning, estadística y NLP.



**Alberto  
Ezpondaburu**

*NLP Specialist*

Alberto es ingeniero de telecomunicaciones, matemático y trabaja de natural language processing engineer en Lang.ai. Es experto en NLP y en la aplicación de técnicas de inteligencia artificial en diversas ramas.



**Luis  
Leite**

*Business Intelligence Consultant*

Experto en la transformación de los datos con experiencia en diferentes áreas como las telecomunicaciones o industria farmacéutica. Posee un master en Física aplicada y otro en Big Data.

# Información **GENERAL**



## **La importancia del Big Data**

Las empresas y organismos oficiales ya se están adaptando a los nuevos tiempos en los que, cuantías de información masiva, se procesan y analizan de tal forma que sean de gran utilidad en multitud de ámbitos diferentes. Lo que hace unos años comenzó siendo una ventaja competitiva de unos pocos, ahora está muy presente y, en breve, será imprescindible para no quedarse atrás.

Por poner unos pocos ejemplos, los macrodatos son una valiosa herramienta en la creación de informes estadísticos complejos, la identificación de nuevas oportunidades de negocio de una empresa, modelos de predicción sobre los resultados publicitarios de una campaña futura o la evaluación de datos masivos para avanzar en investigaciones médicas que ayuden a erradicar enfermedades.

Solo una muestra de como la información a gran escala no afecta únicamente al campo matemático o estadístico. Sus aplicaciones prácticas abarcan todo tipo de entornos reales siendo fundamentales para afrontar situaciones muy diversas a nivel empresarial, gubernamental, científico o social.

## **Salidas profesionales**

- **Gestor de infraestructuras para Big Data**
- **Auditor de sistemas de datos masivos**
- **Arquitecto de Inteligencia de Datos**
- **Arquitecto de Business Intelligence**
- **Chief Data Officer**
- **Data Analyst**
- **Data Consultant**
- **Data Scientist**







#### **1. Preinscripción**

Envía tu solicitud y la documentación para iniciar el proceso.

#### **2. Entrevista**

Dadas las circunstancias, la entrevista se llevará a cabo a través de una videollamada con el Director.

#### **3. Admisión**

Confirmación de la admisión como alumno del Máster Big Data y Business Analytics UCM.

#### **4. Reserva de plaza**

Formalización de tu reserva de plaza como alumno.

## Información DE ADMISIÓN

Preinscribirse cumplimentando el formulario ubicado en la pestaña "Preinscripción". Enviar la documentación requerida a fin de evaluar la candidatura. Entrevista con el solicitante. Confirmación de selección. Formalización de la reserva de plaza.

Tanto la preinscripción como la prematrícula quedan abiertas hasta comenzar el curso académico o completar plazas, estableciéndose lista de espera si procede. Los admitidos deberán ingresar 500 euros en concepto de reserva de plaza para el máster presencial y 350 euros en concepto de reserva de plaza para el máster online. Estas cantidades serán descontadas del importe total de la matrícula. En ningún caso se tendrá derecho a devolución de este importe, a excepción de que no se llegara a celebrar el curso.



### Documentación requerida

- Fotocopia del DNI/pasaporte.
- Certificado de notas oficial.
- Título universitario o resguardo de solicitud de título.
- Currículum Vitae.



### Alumnos con titulación fuera de España

#### **Unión Europea:**

Tanto el título como el certificado de notas tienen que ir acompañados de una traducción jurada.

#### **Fuera de la Unión Europea:**

El título y el certificado de notas tienen que estar legalizados con la Apostilla de la Haya. Si el título y el certificado de notas están en otro idioma que no sea español deben ir acompañados de una traducción jurada.

Una vez finalizados y superados estos estudios, la Universidad Complutense de Madrid emitirá el título, conforme a las normas de admisión y matriculación de los títulos propios de la UCM.



## Información General

# PRESENCIAL



### Horarios y Fechas

**Inicio:** Febrero de 2022

**Fin:** Diciembre de 2022

**Viernes:** De 16:00 a 21:00 h

**Sábados:** De 09:00 a 14:00 h



### Lugar y Créditos

**Lugar:** Facultad de Estudios Estadísticos

**Créditos:** 60 ECTS



### Precio

**Precio:** 6.500€ + 40€ de tasas de secretaría

Pregunta por nuestras becas, facilidades de pago, prácticas en empresas y bolsa de trabajo.

## Procedimiento evaluación

La evaluación de los alumnos se realizará a lo largo de todo el programa a través de ejercicios y casos prácticos.

A la finalización del programa, deberán presentar un Trabajo de Fin de Máster (TFM).

## Materiales e instalaciones

Los alumnos contarán con acceso a una plataforma virtual en la que se encontrará disponible toda la información y documentación relativa al Máster.

El curso se impartirá en aulas de la Universidad Complutense de Madrid, en la Facultad de de Estudios Estadísticos.



Una vez finalizados y superados estos estudios, la Universidad Complutense de Madrid emitirá el título, conforme a las normas de admisión y matriculación de los títulos propios de la UCM.



## Información General

# ON-LINE



### Horarios y Fechas

**Inicio:** Febrero de 2022

**Fin:** Diciembre de 2022



### Lugar y Créditos

**Créditos:** 60 ECTS



### Precio

**Precio:** 4.350€ + 40€ de tasas de secretaría

## Metodología 100% On-line

La formación se realizará de forma tutorizada por los profesores. Se utilizará una plataforma de formación virtual para la comunicación entre los alumnos y profesores, creando una comunidad virtual de trabajo. Los distintos profesores de cada módulo, guiarán a los alumnos proponiendo actividades adicionales dependiendo del temario que se esté cubriendo en cada momento.

## Características plataforma On-line

La plataforma actuará como vía de comunicación entre el alumno y el entorno global de formación.

El estudiante tendrá información actualizada sobre los conceptos que se estén estudiando en cada momento, como enlaces a contenidos adicionales incluyendo noticias, artículos, etc.

Los alumnos deberán realizar y aprobar todas las prácticas de los distintos módulos, y realizar el trabajo fin de máster para poder aprobar el Máster.

### La plataforma cuenta con:

- Mensajería individualizada para cada alumno integrada en la plataforma
- Vídeos
- Videotutorías
- Documentación
- Comunicación con los profesores vía mensajería, foro y chat

Una vez finalizado y superado el Máster de Big Data y Business Analytics, la UCM otorga un título propio, que se rige por las normas de admisión y matriculación de esta Universidad.

Abierto plazo de preinscripción | **PLAZAS LIMITADAS**

**Inicio del Máster:** Febrero de 2022

## CONTACTO



**Teléfono**

[+34 687 30 04 04](tel:+34687300404)



**E-mail**

[info@masterbigdataucm.com](mailto:info@masterbigdataucm.com)

**Sitio Web**

<https://www.masterbigdataucm.com/>