



UNIVERSIDAD
COMPLUTENSE
MADRID

nticmaster

Máster de Formación Permanente en

DATA SCIENCE, BIG DATA & BUSINESS ANALYTICS

7^a
EDICIÓN

Facultad de Comercio y Turismo UCM
Universidad Complutense de Madrid

PROGRAMA

- Módulo . Bases de Datos SQL
- Módulo . Programación Python
- Módulo . Git y Github
- Módulo . Bases de Datos NoSQL
- Módulo . Business Intelligence con Tableau
- Módulo . Estadística
- Módulo . Tecnologías del Big Data
- Módulo . Spark
- Módulo . Minería de Datos y Modelización Predictiva
- Módulo . Machine Learning
- Módulo . Inteligencia Artificial con Deep Learning
- Módulo . Text Mining
- Módulo . Redes Sociales y Big Data
- Módulo . Visualización Avanzada
- Módulo . Productivizar un modelo
- Módulo . Open Data
- TFM . Trabajo Final de Máster



La importancia de la Ciencia de Datos

NECESIDAD DEL MERCADO

Las empresas y organismos están adoptando rápidamente la **transformación digital** para aprovechar el valor de los **datos masivos** en la toma de decisiones. El Big Data se ha convertido en un recurso valioso para la gestión empresarial, y lo que comenzó como una ventaja competitiva se ha vuelto esencial para mantenerse relevante. El dato es considerado como el petróleo del siglo XXI, y aquellos que puedan aprovechar su potencial estarán en una posición privilegiada para **innovar y liderar** en sus respectivos mercados.

APLICACIONES PRÁCTICAS

El impacto de la información a gran escala va más allá del ámbito matemático o estadístico, ya que tiene aplicaciones prácticas en diversos campos empresariales, gubernamentales, científicos y sociales. El Big Data se ha convertido en un recurso fundamental para **afrentar situaciones complejas en tiempo real**, permitiendo tomar decisiones informadas y generar conocimiento valioso en áreas como la medicina, la seguridad, el marketing, entre otros. El Big Data es una herramienta imprescindible en el mundo actual que ayuda a resolver problemas y aprovechar oportunidades en diversos entornos y sectores.

INNOVACIÓN IMPULSADA POR DATOS

El Data Science no solo ayuda a las empresas a ser más eficientes, sino que también impulsa la innovación. Al analizar grandes volúmenes de datos, las organizaciones pueden **descubrir patrones ocultos, nuevas oportunidades de negocio y mejorar procesos**. Esto les permite crear productos y servicios más competitivos y alineados con las demandas del mercado actual.

MACRODATOS

El Data Science y el Big Data son herramientas valiosas en diversas áreas, desde la creación de informes estadísticos hasta la identificación de nuevas oportunidades de negocio, pasando por la predicción de resultados publicitarios y la investigación médica. El análisis de grandes volúmenes de datos permite obtener información relevante y tomar **decisiones informadas en tiempo real**, lo que resulta en una ventaja competitiva para las empresas y avances significativos en la investigación científica.

MEJORA EN LA EXPERIENCIA DEL CLIENTE

Las herramientas de ciencia de datos permiten a las organizaciones personalizar la experiencia del cliente de manera más precisa. Al analizar patrones de comportamiento y preferencias de los usuarios, las empresas pueden ofrecer **productos y servicios más alineados con las expectativas de sus consumidores**, mejorando la satisfacción y la lealtad.



¿Por qué estudiar un Máster de Data Science, Big Data & Business Analytics?

Dentro del mundo empresarial, el análisis de grandes volúmenes de datos se ha vuelto indispensable para la toma de decisiones estratégicas. El Big Data permite a las organizaciones transformar datos en información valiosa para predecir tendencias, optimizar procesos y mejorar el rendimiento. Estudiar un máster en Big Data te preparará para dominar estas técnicas y herramientas, y te permitirá convertirte en un profesional clave en un campo de alta demanda.

Alta Demanda Laboral

Un campo en constante expansión. Las empresas de todos los sectores buscan profesionales capaces de gestionar, analizar y transformar grandes volúmenes de datos en información valiosa para la toma de decisiones. Especializarte en Ciencia de datos te posicionará en un mercado con alta demanda y numerosas oportunidades.

Oportunidades Profesionales

Las salidas profesionales son muy diversas, abarcando roles como analista de datos, ingeniero de datos, arquitecto de datos y científico de datos, entre otros. Un máster te preparará para enfrentar estos desafíos y desarrollar habilidades técnicas avanzadas.

Networking

Estudiar un máster te brindará la oportunidad de conectar con profesionales del sector, compañeros de clase, profesores expertos y empresas. Este entorno te permitirá generar relaciones clave para tu desarrollo profesional y acceder a oportunidades laborales.

Mayor Potencial de Ingresos

Los científicos de datos suelen tener un alto potencial de ingresos debido a la demanda de personas con habilidades técnicas avanzadas. Las empresas están dispuestas a pagar más por aquellos que puedan transformar datos en información estratégica que impulse la toma de decisiones.

Actualización Constante

Un campo que está en continua evolución con nuevas herramientas, tecnologías y enfoques. Un máster te permitirá mantenerte actualizado en las últimas novedades y aprender a adaptarte rápidamente a las tendencias emergentes en la industria de los datos.

Desarrollo de Proyectos Reales

Esta máster te ofrecerá la oportunidad de trabajar en proyectos reales, aplicando tus conocimientos a situaciones del mundo empresarial. Estas experiencias prácticas serán clave para fortalecer tus habilidades y mejorar tu empleabilidad.

Ventaja Competitiva

Tener un máster en ciencia de datos te dará una ventaja competitiva en el mercado laboral, diferenciándote como un profesional preparado para abordar los complejos desafíos de la gestión y análisis de grandes volúmenes de datos.

Flexibilidad Laboral

El conocimiento en Data Science ofrece flexibilidad en el tipo de empresas para las que puedes trabajar, desde startups hasta grandes corporaciones. Además, muchos roles relacionados con Big Data permiten trabajar de forma remota, lo que te da mayor control sobre tu vida profesional.

Emprendimiento

Con los conocimientos adquiridos en un máster en Data Science y Big Data, tendrás las herramientas necesarias para lanzar proyectos propios en el ámbito de la tecnología y los datos, lo que te permitirá explorar nuevas oportunidades de negocio.

Duración

1 año académico

Modalidades

Presencial, Semipresencial y Online

Creditos ECTS

60

Modalidad Presencial

Viernes tarde y sábados
mañana en la universidad

Modalidad Semipresencial

3 Semanas en presencial
en la universidad y en la
plataforma online

Modalidad Online

100% desde nuestra
plataforma online



¿Por qué estudiar en la Universidad Complutense de Madrid?

La Universidad Complutense de Madrid (UCM) es una de las instituciones educativas más destacadas de Europa, reconocida por el prestigioso QS World Ranking como la mejor de España. Ofrece una amplia gama de oportunidades y beneficios para los estudiantes, así como una excelencia académica reconocida, una calidad docente de primer nivel. Ofrece alrededor de 90 títulos de grado y más de 30 dobles grados, más de 200 programas máster, 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 Máster en Formación Permanente?



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. Además, todos los módulos se evalúan con tareas prácticas, no con exámenes, tratándose de un programa de configuración eminentemente práctica.

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.



PROGRAMA

Los módulos de aprendizaje de Data Science, Big Data y Business Analytics más completos enseñados de forma eficaz para el alumno.



UNIVERSIDAD
COMPLUTENSE
MADRID

 **nticmaster**



MÓDULO

Bases de Datos Relacionales

Diseño y modelización de base de datos y lenguaje de consulta estructurada. Estudio del modelo entidad-relación, modelo relacional, e implementación relacional con SQL.

En éste módulo de base de datos relacionales, se estudiarán los fundamentos y conceptos necesarios de los sistemas de gestión de bases de datos relacionales y cómo se pueden implementar y administrar utilizando el lenguaje de consulta estructurado.

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 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.

Índice de contenidos:

- 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. Transformación del modelo entidad/relación a tablas.
- **Conceptos fundamentales del modelo relacional** como relaciones, claves y restricciones de integridad, así como normalización.
- **Implementación relacional:** lenguaje de definición de datos (DDL) y lenguaje de manipulación de datos (DML). Consultas simples, subconsultas, JOIN, EXIST y NOT EXIST, HAVING Y GROUP BY, UNION, INTERSECT Y EXCEPT, INSERT, UPDATE Y DELETE. Vistas, trigger, procedimientos almacenados y funciones.



MÓDULO

Programación Python

Python es el lenguaje de programación más utilizados en la actualidad en el ámbito de Big Data y Data Science, ya que es un **lenguaje intuitivo y con sintaxis clara y sencilla**, con amplia gama de librerías y compatibilidad con otras tecnologías, y con una gran comunidad activa de desarrolladores.

En este módulo adquiriremos los conocimientos necesarios para programar utilizando el lenguaje Python, de forma escalonada y mediante ejercicios prácticos que complementen la teoría, seremos capaces de programar modularmente en dicho lenguaje haciendo uso de su sintaxis y resolviendo los supuestos planteados en clase. Los conocimientos adquiridos nos proporcionarán las herramientas necesarias para afrontar el conocimiento de Python orientado al entorno **big data y machine learning**.

Al final del módulo se introducirán algunos conceptos básicos y librerías utilizadas en el tratamiento de datos y cálculo numérico como son **Pandas y NumPy**, así como desarrollo de una API que sea capaz de exponer datos y un programa que sea capaz de ingestar datos de una API.

Índice de contenidos:

- Variables.
- Tipos y operadores.
- Entrada/Salida de Datos.
- Estructuras de Control I (IF Else, While, For, Break / continue).
- Listas.

- Tuplas.
- Diccionarios.
- Estructuras de Control II (For).
- Funciones.
- Módulos.
- Plataformas de desarrollo.
- PyCharm.
- Métodos de las colecciones, cadenas de texto, listas, conjuntos y diccionarios.
- Módulos.
- Ficheros y objetos con pickle.
- Pandas.
- NumPy.
- Bases de datos.
- MongoDB.
- SQL Server.
- Twitter como fuente de datos.
- Escribir/Leer en un fichero.



```
0 response = requests.get(url)
1 # checking response.status_code (if you get 502, try refreshing the page)
2 if response.status_code != 200:
3     print(f"Status: {response.status_code} - Try refreshing the page!")
4 else:
5     print(f"Status: {response.status_code}")
6
7 # using BeautifulSoup to parse the response object
8 soup = BeautifulSoup(response.content, "html.parser")
9
10 # finding Post images in the soup
11 images = soup.find_all("img", attrs={"alt": "Post image"})
12
13 # downloading images
14 for image in images:
15     # ...
```



MÓDULO

Git y Github

En este módulo se trabajará el entorno de trabajo con Git, conocimiento esencial para preparar a los estudiantes en el uso de herramientas fundamentales para la gestión de proyectos y el análisis de datos a gran escala.

Git se utiliza ampliamente en la industria de Big Data debido a su flexibilidad, escalabilidad y capacidad para manejar grandes cantidades de datos. A través de este módulo, los estudiantes aprenderán a realizar tareas avanzadas de programación y análisis de datos.

Git es un sistema de control de versiones que se utiliza para administrar el código fuente de los proyectos de software. Los estudiantes aprenderán cómo usar Git para colaborar en equipo, compartir código y hacer un seguimiento de los cambios en el código a medida que se realiza la edición y mejora del software.

Índice de contenidos:

- Git
- GitHub
 - Creación de repositorios
 - Documentación del proyecto (Markdown)
- Instalación y configuración de Git
- Comandos básicos en Git
- GitFlow - Trabajando con ramas



MÓDULO

Bases de Datos NoSQL

En éste módulo 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.

Índice de Contenidos

Introducción a las bases de datos NoSQL

- Un poco de historia sobre las BD relacionales.
- Modelo de datos (entidad relación vs agregación).
- BD Relacionales VS NoSQL (ACID vs BASE).
- Diferentes tipos de BD NoSQL (key-value, documentales, XML, Grafos, Column Family).
- Cuando utilizar NoSQL.
- Teorema del CAP.
- Mongo VS Cassandra.

Instalar MongoDB

- Cómo instalar servidor MongoDB en WIN/MAC/Linux.
- Cómo conectarse a mongo con diferentes clientes.
- Conceptos de JSON y Arrays.
- Manejo básico en la base de datos/colecciones y tipos de datos.
- Importación y Exportación de ficheros.
- Operaciones CRUD.

Create – insert, insertOne, insertMany

- Read – find, findOne.
- Update – update, updateOne, updateMany, findAndModify y operadores.
- Delete – remove, deleteOne, deleteMany.

Dominar el Find

- Aplicar filtros avanzados.
- Arrays y Documentos Embebidos y operadores.
- Ordenar los resultados de búsqueda.
- Paginar y limitar los resultados de búsqueda.
- Otros cursores importantes.

Índices y Agregaciones

- Índices.
- Agregaciones, pipeline, etapas, operadores.

Modelo de datos

- Cómo modelar.
- Denormalización.

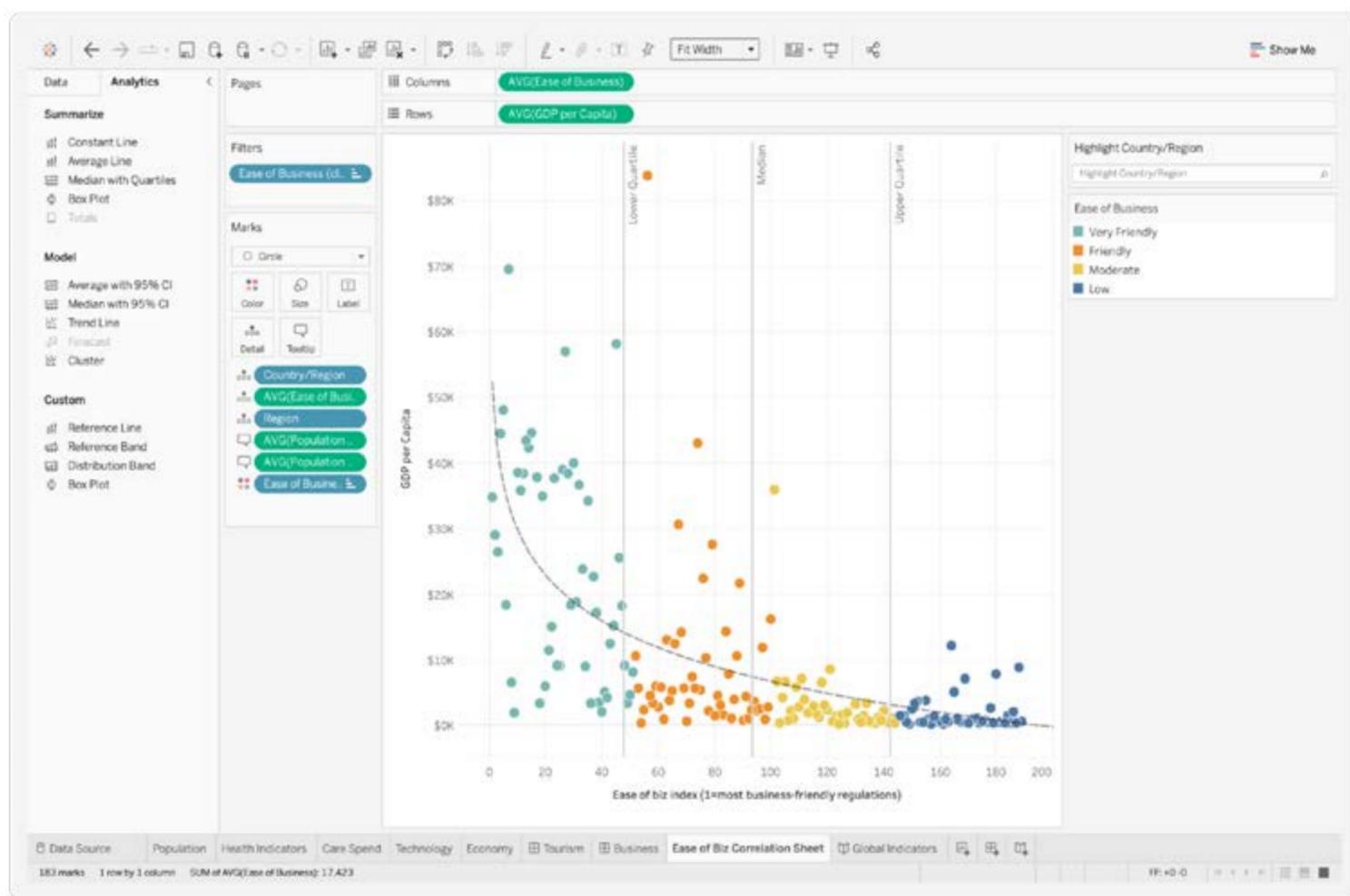
MÓDULO

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 Machine Learning o de Data Science, 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 años. 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 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

Fundamentos de estadística

La estadística permite **sistematizar la información** proporcionada por las bases de datos cuya procedencia, como sabemos, puede ser muy variada: datos económicos de empresas, datos sobre la realidad social, datos sobre la economía, sobre deportes, finanzas, opinión pública y política, etc. El tratamiento estadístico de los datos, con frecuencia, no acaba en la obtención de un resumen de la información (Estadística Descriptiva), sino que sirve para **tomar decisiones en múltiples ámbitos, resolver diversos tipos de problemas y generar modelos predictivos**.

Gracias a este módulo los alumnos serán capaces de manejar las herramientas que les permiten la estimación de los parámetros de los modelos, los contrastes de hipótesis sobre dichos parámetros, así como el acceso a la metodología que permite el rigor en la toma de decisiones en Estadística.

Se persigue la consolidación de los conocimientos de estadística necesarios para abordar el resto de las áreas del máster.

El módulo está dividido en dos temas:

Estadística descriptiva:

Se centra en la descripción de las **variables estadísticas univariantes**, que estudian 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**.

Algunos de los puntos que se tratarán son:

- 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 estadística:

Cuando hablamos de inferencia en el contexto estadístico, nos estamos refiriendo a las diversas técnicas y metodologías a través de las cuales, en base a una información, los datos, con la que previamente contamos, se realizan **modelos de predicción**. Además es muy relevante controlar además el **margen de error** en las decisiones, margen que necesariamente existe dada la aleatoriedad intrínseca en los fenómenos que analiza la estadística.

Un ejemplo muy cotidiano al respecto son las encuestas electorales de intención de voto, donde con 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. Pero también el análisis del riesgo a la hora de conceder un crédito personal o hipotecario, el uso de datos para el diagnóstico diferencial en medicina preventiva, o la información que recibe un portero de fútbol sobre las habilidades y tendencias del delantero que le tirará el penalti.

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

Tecnologías del Big Data

El alumno aprenderá a descubrir y valorar oportunidades presentes en el día a día del ciudadano y de la empresa.

Se verán temas como el **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. También se tratará la aproximación a la aportación del IOT al mundo del macrodato y la conectividad con fuentes de datos heterogéneas a través de **bróker de mensajes** y **hubs** con dispositivos

Otros temas a tratar serán:

- 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.

Índice de contenidos:

Hadoop:

- Introducción a Hadoop, entorno de trabajo que soporta diversas aplicaciones distribuidas bajo una licencia libre.
- Almacenamiento HDFS para grandes volúmenes de datos en Apache Hadoop.
- Acceder a los datos almacenados en HDFS como si fuera una base de datos: introducción a HIVE.

Spark:

- Introducción a Spark, sistema de computación basado en el paradigma de Hadoop que destaca por su velocidad gracias al almacenamiento de datos en memoria.
- Clúster de Spark.
- Desarrollo de aplicaciones con Spark.
- RDD, Datasets y transformaciones.
- Spark Streaming.

Tecnologías Big Data en Google Cloud:

- Ingesta y procesamiento en batch: Cloud Dataproc.
- Ingesta en streaming: Cloud PubSub.
- Almacenamiento de ficheros: Google Cloud Storage.
- Almacenamiento de datos: BigQuery.
- Desarrollo y ejecución de pipelines: Cloud Data Fusion.
- Visualización de datos: Cloud Data Studio.
- Analítica avanzada de datos: Vertex AI.
- Orquestación de datos: Cloud Composer.
- Catálogo y gobierno del dato: Cloud Dataplex.



MÓDULO

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. Proporciona a los estudiantes una comprensión profunda de cómo funcionan estos **sistemas de procesamiento de datos distribuidos** y cómo aprovecharlos para procesar grandes cantidades de datos de manera eficiente y efectiva.

Tras describir brevemente el manejo de **HDFS**, el curso se centrará en **Apache Spark**, sin duda la tecnología más demandada para procesamiento de grandes volúmenes de datos. Describiremos su filosofía peculiar basada en un grafo de ejecución (DAG) y profundizaremos en cada uno de los módulos, en especial Spark SQL, MLlib y Structured Streaming. Se usará el servicio **Dataproc** de Google Cloud para que cada alumno pueda desplegar un cluster de forma sencilla y gratuita.



Índice de contenidos:

- Introducción a las tecnologías Big Data.
- HDFS.
- Apache Spark.
- Arquitectura de Spark.
- Spark SQL.
- Spark MLlib.
- Spark Streaming.
- GraphFrames.



MÓDULO

Minería de Datos y Modelización Predictiva

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, 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.

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.

Índice de contenidos:

- **Depuración de datos y modelos de regresión:** esta primera parte se centra en la preparación del conjunto de datos de cara a la modelización predictiva de tipo supervisada y la presentación de los modelos clásicos de predicción para variables continuas mediante regresión lineal y variable dicotómica mediante regresión logística.

- **Series temporales y modelos no supervisados:** en la segunda parte se abordan los métodos de estudio y predicción de series temporales univariantes clásicas como introducción a los estudios temporales y las dos principales técnicas no supervisadas (no hay variable objetivo específica) para reducción de dimensiones y segmentación de los datos.
- Modelos de puntuación de **Riesgo de Crédito** (Scoring de riesgo de crédito).
- **GIS y Modelos predictivos** con Datos Espaciales.





MÓDULO

Machine Learning

El objetivo principal del módulo es enseñar a los estudiantes cómo **diseñar, implementar y evaluar modelos de Machine Learning** que puedan aprender de los datos y hacer predicciones o tomar decisiones. Los alumnos aprenderán diversas técnicas 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.

Entre las aplicaciones actuales encontramos detección de fraudes, el análisis de sentimientos, la clasificación de imágenes, reconocimiento de voz e imagen, automatización, cuidado de la salud, entre otros.

Índice de Contenidos

Se estudiarán los conceptos básicos necesarios para entender e implementar métodos de **Aprendizaje Automático o Machine Learning** y realizaremos un repaso general de las principales **librerías** usadas en el desarrollo de algoritmos de Machine Learning.

Seguidamente, realizaremos la **Ingesta de Datos**, se explorarán distintos tipos y fuentes de datos. Además, se estudiarán casos reales de obtención de datos y como convertirlos en **Data Frames**, y se aprenderá a emplear mecanismos de organización, estructuración, análisis y visualización de los datos con ejemplos prácticos.

Continuamos con la fase de **preprocesado**, se estudiarán y aplicarán distintas técnicas de procesado de datos para eliminar el ruido en ellos, mejorar su calidad, su capacidad representativa, o aumentar su cantidad y variedad. Entre las técnicas que se tratarán se encuentra el filtrado, balanceo y una amplia variedad de métodos de aumento de datos.

Se aprenderán y practicarán las características generales y particulares de los modelos **"Machine Learning"**, centrándose en los algoritmos supervisados de regresión y principales clasificadores (Naive Bayes, SVM, Random Forest y Logistic Regression), así como los algoritmos no supervisados de clustering (PCA, K-Means), sus características y aplicaciones a datos como iris, dígitos, textos, imágenes, etc.

Además, conoceremos **diferentes herramientas** que nos ayudarán en el día a día en nuestro trabajo como Data Scientist como: autoML (pycaret) para modelado, EDA (pandas profiling), explicabilidad (lime) para explicar lo que ha ocurrido en el entrenamiento del modelo, mlflow para monitorizar métricas, h2o con cluster, feature engineering para trameo de variables, Ensemble algoritmos de alto rendimiento, etc.

El programa no se detendrá únicamente en 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. A través de varios ejemplos prácticos, analizaremos cómo funciona la industria y cómo enfrentarnos a diferentes casuísticas.





MÓDULO

Inteligencia Artificial con Deep Learning

El contenido de la sección de deep learning estará estructurado en **cuatro bloques**. 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 un **componente práctico** en cada bloque para familiarizar a los alumnos con la aplicación directa.

Índice de contenidos:

1. Redes Neuronales:

En este bloque se pretende introducir el concepto de red, explicando su funcionamiento y permitiendo que el alumno lo comprenda 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.

2. Redes Convolucionales. Imágenes:

En este segundo bloque se presentará una de las estructuras 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. Los contenidos de este bloque son los siguientes:

- 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.

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. Los contenidos de este bloque son los siguientes:

- 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.

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. Los contenidos de este bloque son los siguientes:

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



MÓDULO

Text Mining

Desde el inicio de la Era de la Información, se ha producido un proceso de **generación de datos**, muchos de los cuales se encuentran en **formato no estructurado**. Estos datos requieren de técnicas de análisis específicas para su procesamiento.

En este bloque, los estudiantes adquirirán conocimientos fundamentales sobre Text Mining, una subárea de la ciencia de datos que posibilita la **transformación de los datos textuales no estructurados a un formato tabular para su análisis**, identificación de patrones o extracción de conocimiento.

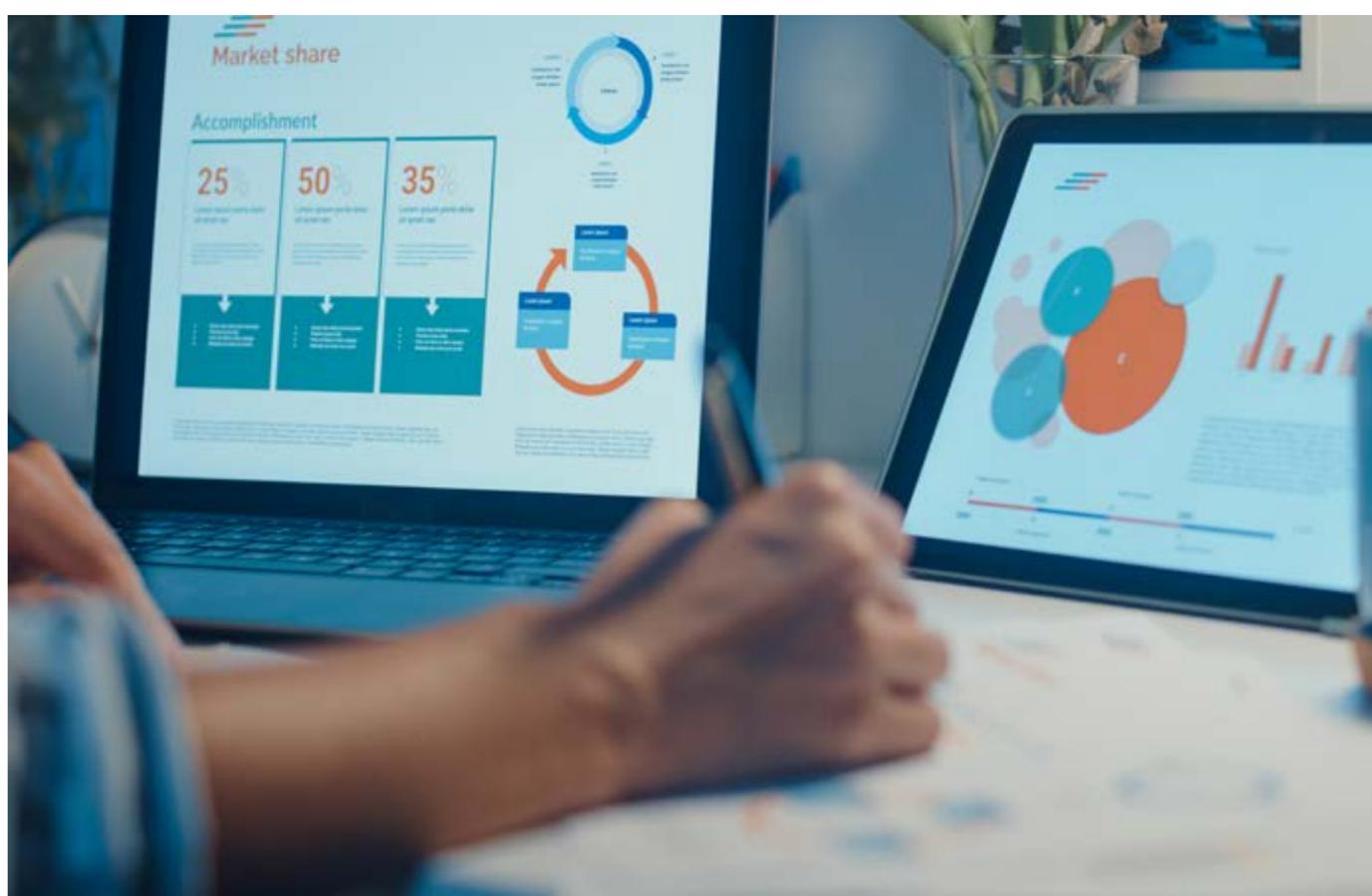
En este módulo se expondrá la relevancia de las **tecnologías del lenguaje** para el análisis de información textual. Se examinarán las distintas técnicas de análisis de textos existentes, los fundamentos del Procesado de Lenguaje Natural (PLN) y las diversas maneras de representar los textos de forma numérica para su procesamiento por medio de algoritmos de inteligencia artificial. Se abordarán temas como el preprocesamiento de textos (limpieza, transformación, obtención de raíces, etc.), su exploración y su procesamiento

con aprendizaje automático.

En el módulo se abordarán dos problemas clásicos de Text Mining. El primero de ellos es la **clasificación de textos**, que permite la generación de modelos de aprendizaje automático capaces de asignar etiquetas a los textos. El segundo es el **Topic Modeling**, una técnica utilizada para analizar el contenido de grandes volúmenes de documentos.

Además, se aplicarán técnicas de clasificación en un caso de estudio sobre la extracción de sentimientos de textos provenientes de redes sociales. El Text Mining se encuentra en un momento de máximo desarrollo, en el que se han desarrollado notables tecnologías como **ChatGPT**, construidas sobre arquitecturas de redes neuronales profundas.

En este módulo, también se presentarán las tecnologías que han facilitado el rápido desarrollo de la técnica como son los modelos de lenguaje y las **redes neuronales de tipo Transformers**. En el desarrollo de este bloque se utilizará el lenguaje Python y librerías como NLTK, Spacy, Gensim o Transformers.





MÓDULO

Redes Sociales y Big Data

El análisis de redes sociales busca **representar relaciones entre distintas entidades** para poder describir la estructura de la red social o económica resultante. Esta forma de análisis nos permite estudiar múltiples conjuntos de datos que no tienen cabida su análisis en el marco de una Dataframe.

Algunos ejemplos son el análisis de la información entre las páginas Web, en las Redes Sociales de personas (Facebook, Instagram, Twitter, LinkedIn, TikTok...), en las Apps de comunicación (Whatsapp, Facebook Messenger, Skype...) y en general cualquier entorno donde dispongamos de **información relacional entre los usuarios** que deseamos transformar en conocimiento y valor de negocio (relaciones entre clientes, relación entre suministros, relación entre trabajadores, relaciones entre publicaciones, relaciones entre países...)

Los principales análisis que se desarrollan se deben englobar en tres grandes grupos. En primer lugar, el **análisis estructural de la red**, imprescindible para avanzar en el análisis, y punto de partida para predecir el comportamiento futuro de las relaciones modeladas. En segundo lugar, el **análisis de la centralidad o la importancia de los sujetos** de la Red, objetivo en sí de muchos análisis y dado inevitable para dar valor al negocio que está detrás de los datos de la Red. Por último, la **identificación de comunidades** en la Red, herramienta esencial para explicar la dinámica de comunicaciones en la Red y abordar situaciones como la identificación.





MÓDULO

Visualización avanzada

El módulo de Visualización Avanzada es una parte fundamental en el análisis de datos en el mundo de Big Data.

La capacidad de **visualizar datos de manera efectiva y clara** es esencial para la toma de decisiones y para la comunicación de resultados en un entorno empresarial.

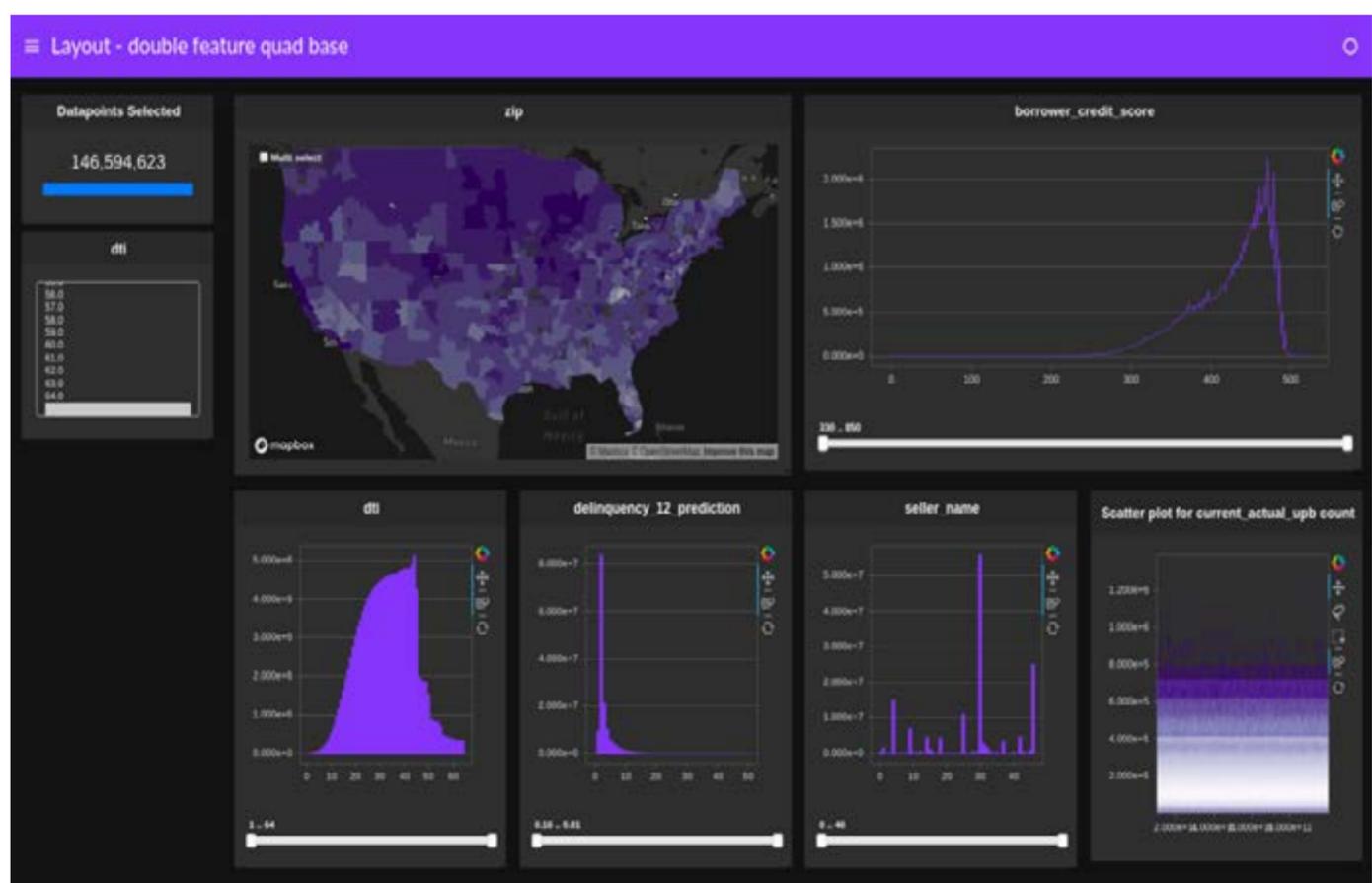
Índice de Contenidos

En el módulo se abordarán los **entornos Python**, como los notebooks y los entornos virtuales, que son comunes en el mundo de Big Data.

También se introducirá el concepto de **gráficos interactivos** y se destacarán los beneficios de la interactividad en la presentación de datos.

Se exploran las herramientas **Bokeh** y **Pygal** y se presentan **dashboards** en Python utilizando Dash y Glean.

También se aprenderán las **opciones de mapas en Python**, incluyendo Geoplotlib y Leaflet/Folium, y también se trabajarán las animaciones como una forma efectiva de presentar datos, así como la librería Plotly para gráficos interactivos.





MÓDULO

Productivizar un modelo

Productivizar un modelo es esencial para aplicar de manera adecuada los conocimientos adquiridos durante los anteriores módulos del máster, así como los modelos de machine learning de manera eficiente y efectiva en el mundo real.

Una vez que se ha desarrollado un modelo, el siguiente paso es **implementarlo en un sistema de producción para que pueda ser utilizado para hacer predicciones o clasificaciones en tiempo real**. El módulo de productivizar un modelo es la herramienta que ayuda a automatizar este proceso de implementación.

Este módulo es responsable de **integrar el modelo en el flujo de trabajo del sistema de producción y asegurar que el modelo pueda funcionar** correctamente en un entorno en vivo. Algunas de las tareas que el módulo de productivizar un modelo realiza son:

- **Preprocesamiento de datos:** el módulo prepara los datos de entrada para el modelo, lo que puede incluir el procesamiento de datos faltantes, la normalización de datos, la selección de características y el muestreo de datos.
- **Integración del modelo:** el módulo integra el modelo de machine learning en el sistema de producción y establece las interfaces necesarias para que el modelo pueda ser utilizado por otros componentes del sistema.
- **Monitoreo del modelo:** el módulo de productivizar un modelo monitorea constantemente el rendimiento del modelo y alerta al equipo de cualquier problema o desviación en los resultados de las predicciones.
- **Escalado y actualización:** el módulo de productivizar un modelo también es responsable de garantizar que el modelo pueda manejar grandes volúmenes de datos y de actualizar el modelo a medida que se reciben nuevos datos para mejorar su precisión.



MÓDULO

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 puede ser Kaggle.

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, sino 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.

Índice de Contenidos

- **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**.





TFM

Trabajo Fin de Máster

Asimilados todos los conceptos previos, llega el momento de poner a prueba los conocimientos adquiridos en el máster.

El trabajo de fin de máster es una **parte crucial del programa**, ya que permite a los estudiantes aplicar todos los conocimientos adquiridos en el curso en un proyecto práctico y relevante en el mundo real.

Proporciona una oportunidad para que los estudiantes demuestren su capacidad para **analizar, procesar y utilizar datos de manera efectiva para resolver problemas complejos y tomar decisiones informadas**.

También les permite desarrollar habilidades de presentación y comunicación al presentar sus hallazgos y resultados a una audiencia de expertos.

Además, el trabajo de fin de máster puede ser una oportunidad para que los estudiantes trabajen en **colaboración con empresas u organizaciones**, lo que les permite obtener experiencia práctica en un entorno profesional y crear conexiones valiosas para su carrera.



EQUIPO DOCENTE

*Clases y tutorías con grandes
profesionales del sector de la
Ciencia de Datos y el Big Data.*



UNIVERSIDAD
COMPLUTENSE
MADRID

 **nticmaster**



EQUIPO DOCENTE

Directivos



María
Concepción García

Vicerrectora de Empleabilidad y Emprendimiento UCM Departamento de Organización de Empresas y Marketing UCM. Profesora del Dpto. de Organización de empresas. Lic. en Ciencias Económicas y Empresariales.



Jose Carlos
Soto Gómez

Co-Director del Máster. 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...

Coordinadores



David
del Ser

Coordinador Máster

David es Lic. en Marketing por ESIC, Honours Degree in Business Administration por Humberstone University, MBA por UNED, Máster Dirección Financiera, Máster Marketing Digital, Máster en Big Data. Especialista en el desarrollo de negocio y transformación digital en Ntic Master. Gran experiencia profesional trabajando en Grupo Iberostar, Grupo Avintia, entre otras.



Cristóbal
Martínez Martínez

Coordinador Máster

Cristóbal es Ingeniero informático. Director de IT en Aplimovil y Ntic Master. Profesor máster marketing digital de la UCM, UNED, Cámara de Comercio y CEEIC. Experto en sistemas y procesos informáticos. Gran experiencia profesional trabajando en empresas referentes como NEC, BNP Paribas, Banco de España, Vodafone.



EQUIPO DOCENTE

Docentes



María José
Gómez Silva
Docente UCM

María José es docente en la UCM y Doctora en Ingeniería Eléctrica, Electrónica y Automática e Ingeniera Técnica Industrial. Experta en machine learning e Ingeniería Eléctrica, Electrónica y Automática.



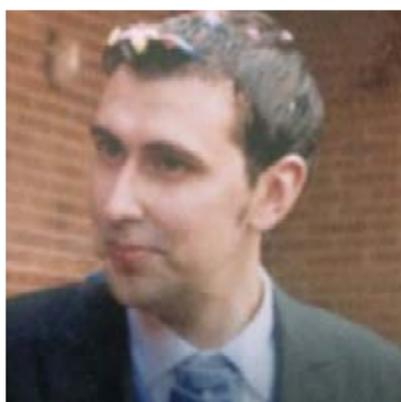
Lorenzo
Escot Mangas
Profesor Titular de la UCM

Lorenzo es Doctor en Economía. Experiencia de más de 20 años como docente y Codirector del Grupo de Investigación Análisis Económico de la Diversidad. Experto en análisis de Datos en Economía Aplicada, Econometría Aplicada y Análisis de datos Espaciales.



Gabriel
Marín Díaz
PhD, Análisis de Datos

Más de 25 años de experiencia en la mejora de procesos y definición de modelos de negocio. Especialista en Analítica de Datos, Customer Experience, Consultoría Tecnológica y Gestión de Equipos.



José Javier
Galán Hernández
Responsable de Sistemas

José Javier es Ingeniero Informático y trabaja como Responsable de Sistemas CED. Además es profesor asociado UCM. Ha trabajado en proyectos de sistemas en El Corte Inglés y Comel, entre otros.



Armando
Heras
Chief Digital Officer

Armando es Licenciado en Economía por CUNEF. Experto en finanzas, auditorías y financiación, especialmente en ecosistemas emprendedores a través de la inversión, creación de empresas y desarrollo de proyectos.



Eduardo
Fernández Carrión
Data Scientist, PhD

Eduardo es ingeniero informático y doctor en métodos estadísticos matemáticos para el tratamiento computacional de la información. Data Scientist & ML Engineer en Santalucia seguros. Experto Data Scientist habiendo desarrollado su carrera profesional en StratioBD, Visavet, Everis, etc.



EQUIPO DOCENTE

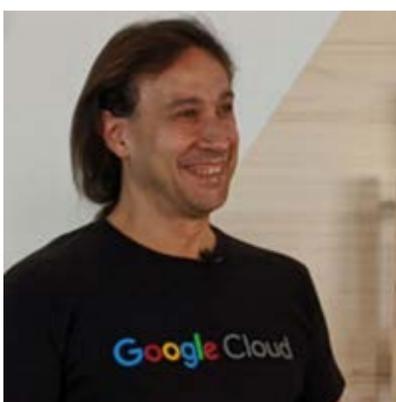
Docentes



Santiago
Mota Herce

Corporate Advisor

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



Pedro Pablo
Malagón Amor

Principal Architect en Google

Pedro Pablo es ingeniero de sistemas. Experto en Big Data, Data Science, Data Engineer. Tiene una gran experiencia profesional de más de 23 años en Microsoft como Principal Architect y actualmente en Google.



Pedro
Concejero Cerezo

Project Manager at Telefónica
Investigación y Desarrollo

Experto en la aplicación de tecnologías Big Data, Machine Learning y Deep Learning en la resolución de problemas de negocio.



Pablo J.
Villacorta

Científico de Datos en
Santalucía Seguros

Pablo es 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.



Luis
Gascó Sánchez

NLP Research Engineer

Senior Machine Learning Engineer en Avature. Doctor en Ingeniería por la Universidad Politécnica de Madrid con una larga trayectoria profesional en proyectos de I+D+i relacionados con IA y NLP. Ha trabajado en centros de investigación de prestigio como Barcelona Supercomputing Center, Nokia Bell Labs (UK) o Télécom ParisTech (Francia), entre otros.



José Ángel
Carballo Sánchez

Data Scientist/ IA Engineer at
Telefónica

José Ángel es ingeniero de Telecomunicaciones. Experto en NLP e IA. Trabaja como Data Scientist & Artificial Intelligence Engineer en Telefónica y ha trabajado Thinl Big Analytics, LLC y proyectos Data Science en Bankia y Vodafone.



EQUIPO DOCENTE

Docentes



Javier Domínguez Gómez

Cryptographer & Cybersecurity Engineer

Javier es Ingeniero informático con certificaciones en CISSp, Cisco, entre otras. Experto en Ciberseguridad y Data Science. Desarrolla su carrera como Cybersecurity Engineer | D&R para BBVA.



Guillermo Villarino Martínez

Doctor en Data Science

Doctorado en Data Science por la UCM. Sus principales líneas de investigación están centradas en modelos de Machine Learning y tratamiento de la información para la toma de decisiones en el ámbito de la Clasificación Supervisada.



Charles Flores Espinoza

Big Data Engineer en Mercedes-Benz AG

Charles es Ingeniero informático. Experto Data Engineer & Data Architect, Scala, ML. Gran experiencia en el sector en empresas como StratioBD, Oliver Wyman, Vass, Bayes.



Carlos Ortega Fernández

Expert Data Scientist

Carlos es Lic. en Físicas y trabaja como Lead Specialist en SDG Group España. Tiene una dilatada experiencia en el sector con aplicaciones avanzadas basadas en datos y Machine Learning en múltiples sectores.



Álvaro Bravo Acosta

Ingeniero Técnico Informático en Sistemas

Experto en Tecnologías Big Data, BI y Analítica. Gran experiencia en consultoras como Minsait, Sopra Steria, Everis o NTT, para externos como ISBAN y BBVA. Actualmente, DevOps en equipo de producto Frameworks en Strato.



Tomás Morla

Ingeniero Técnico Informático en Sistemas

Ingeniero de datos y profesional al mundo del dato. Con años de experiencia, en experimentación, transformación y consulta de información en diferentes entornos tecnológicos. Especialista en la creación de repositorios de datos destinados a que las áreas de negocio pueden obtener valor de él.



EQUIPO DOCENTE

Docentes



Alberto Ezpondaburu
NLP Specialist

Alberto es ingeniero de telecomunicaciones, Lic. en Matemáticas. Experto en NLP y en la aplicación de técnicas de Deep Learning e Inteligencia Artificial. Ha trabajado como Data Scientist en varias empresas y creado empresas del ámbito AI.



Adolfo Hernández Estrada
Profesor Titular de la UCM

Adolfo es Licenciado y Doctor en Ciencias Matemáticas. Director del Departamento de Economía Financiera y Actuarial y Estadística. Director del grupo MECOCISO. Experto en economía y actuariales.



Luis Garmendia
Profesor Titular de la UCM

Luis es Lic. en matemáticas con especialidad en ciencias de la computación. Departamento de Ingeniería del Software e Inteligencia Artificial (DISIA) UCM. Miembro del EUSFLAT – European Society for Fuzzy Logic and Technology. Miembro del Grupo de Soft Computing de la facultad de Informática de la UCM.



Juan Carlos Fernández
Chief Technology Officer
Mirai Advisory

Juan Carlos es ingeniero de telecomunicaciones. Experto Data & Analytics Engineer, Business Analysis, Business Intelligence, and Big Data. Ha trabajado como data scientist en Vodafone, Neoris, entre otras.



Pablo Ruiz
Machine Learning Engineer en
Apple

Licenciado en Ingeniería Química por la Universidad Politécnica de Madrid. Actualmente es Ingeniero de Machine Learning en Apple y ha sido encargado del Machine Learning and Web Development en Twilio. Pablo es experto en Python, Natural Language Processing (NLP).



Javier Castro Cantalejo
Docente en la UCM

Javier es Doctor por la Universidad Complutense de Madrid con la tesis Reparto de holguras y costes en una red Pert 2007. Javier tiene experiencia, como miembro de grupos de investigación. Experto en «Text Mining y Redes Sociales».



EQUIPO DOCENTE

Docentes



Juan Fernando
Sánchez Martínez

Data & Analytics Trainer en Telefónica

Profesor de Data & Analytics experto en Bases de Datos SQL, Business Intelligence e Inteligencia Artificial. Ingeniero Industrial Químico (UPV), posee una sólida experiencia en investigación adquirida en la Universidad de Castilla-La Mancha y en la Universidad La Sapienza di Roma. Colabora con reconocidas empresas como Telefónica, BBVA y SDG Group.



Ana Lucila
Sandoval

Docente UCM

Doctora en Ingeniería Informática por la UCM. Investigadora y experta en la teoría de la codificación, la seguridad de la información y sus aplicaciones.



José Carlos
Sánchez

Analytics Engineer

José Carlos es Graduado en ingeniería química. Experto Data Analyst and Business intelligence. Trabaja en SDG Group como senior qualified Analytics Engineer y ha trabajado como Analytics Engineer para Accenture, Pagantis y otras empresas del sector.



Carlos
Baraña Arias

Specialist Engineer en Ingenia Telecom

Master en Big Data y Business Analytics, tecnología y procesamiento de datos. Ingeniero de telecomunicaciones.



Rubén
López Rodríguez

Ingeniero informático

Ingeniero informático y máster en formación, cuenta con más de 20 años de experiencia como analista programador y docente. Experto en tanto en programación y robótica como en Git.



Tú

Futuro experto en Data Science

Porque con nosotros aprenderás sobre Ciencia de Datos, y aunque el camino tienes que recorrerlo tú, quizás muy pronto estés aquí como profesor nuestro.



EQUIPO DOCENTE

Ponentes Masterclass



José Ignacio
Carrión

Private Equity Mandate Officer
en Fondo de Inversión Europeo

José Ignacio se dedica al desarrollo, implementación y gestión de mandatos de inversión con clientes institucionales públicos. Supervisa y gestiona las inversiones en fondos de fondos privados a nivel europeo.



Guillermo
Campoamor

CEO de Meep

Tras pasar gran parte de su vida laboral viajando por más de 22 países, decidió emprender en el sector de la movilidad. Desde entonces, Meep ha levantado más de 2 millones de euros y ha sido elegida como una de las 10 startups más relevantes en movilidad por #SouthSummit18



Ángel
Araujo

Southern Europe General
Manager

Responsable del negocio de Evaneos para España e Italia. Responsable de la P&L y de un equipo de 35 personas. Diseño y ejecución de las diferentes palancas de la plataforma: Oferta, Demanda y Marketing. . Managing Partner y Co-Founder de Unlock Management.



Armando
Heras

Chief Digital Officer

Armando Heras, Director de la Estrategia digital de CMC XXI, especializado en finanzas, auditorías y financiación. Su función reside, entre otras cosas, en crear ecosistemas emprendedores a través de la inversión, creación de empresas y desarrollo de proyectos.



Matteo
Consonni

Former startup founder

Matteo cuenta con experiencia en desarrollo de negocios, gestión de programas y financiación en 4 países. Actualmente, dirige las actividades de aceleración de Startups para Health Kic del Instituto Europeo de Innovación y Tecnología (EIT)



SALIDAS PROFESIONALES

El sector de Tecnologías de la Información continuará siendo el motor principal del crecimiento laboral, encabezando el top de empleos más demandados y mejor remunerados por las empresas.



UNIVERSIDAD
COMPLUTENSE
MADRID

 nticmaster



Salidas Profesionales

El campo de Data Science, Big Data y Business Analytics ofrece una gran variedad de salidas profesionales debido a su relevancia creciente en la estrategia empresarial moderna. Desde la analítica avanzada hasta la ingeniería de datos a gran escala, estas disciplinas proporcionan oportunidades clave para transformar los datos en soluciones estratégicas. A continuación, te presentamos algunas de las muchas oportunidades profesionales que podrás explorar en este ámbito en constante evolución:

Con el avance de la tecnología y la evolución constante del entorno digital, es probable que sigan surgiendo nuevas especialidades y roles.

Data Analyst

Analiza conjuntos de datos grandes para obtener insights útiles que apoyen las decisiones estratégicas de una empresa.

Auditor de sistemas de datos masivos

Revisa y audita sistemas que gestionan grandes volúmenes de datos para asegurar su precisión, eficiencia y cumplimiento regulatorio.

Arquitecto de Business Intelligence

Diseña y mantiene las soluciones de inteligencia empresarial que permiten a las empresas transformar datos en información procesable.

Desarrollador de datos a gran escala

Especialista en crear infraestructuras que permiten almacenar y procesar datos masivos de forma eficiente.

eCommerce Data Analyst

Analiza datos relacionados con el comercio electrónico, optimizando ventas, estrategias de marketing y experiencia de cliente.

Digital Transformation for Companies

Especialista en implementar tecnologías digitales para optimizar procesos, productos y servicios en una empresa.

Data Steward

Responsable de la gestión y gobernanza de los datos, asegurando su calidad, seguridad y disponibilidad en toda la organización.

Data Artist

Presenta datos complejos de manera visual y comprensible, ayudando a las organizaciones a tomar decisiones basadas en datos a través de visualizaciones atractivas.

Chief Data Officer (CDO)

Ejecutivo responsable de liderar la estrategia de datos en la organización, asegurando su adecuada gestión y aprovechamiento.

Data Scientist

Utiliza machine learning, estadísticas avanzadas y otras herramientas para interpretar datos complejos y resolver problemas empresariales mediante análisis predictivos.



ADMISIONES

Tanto la preinscripción como la pre matrícula quedan abiertas hasta comenzar el curso académico o completar plazas.



UNIVERSIDAD
COMPLUTENSE
MADRID

 nticmaster



Proceso de admisión



Preinscripción

Envío de solicitud para evaluar candidaturas.

1. Preinscribirse cumplimentando el formulario ubicado en la pestaña "Preinscripción" de la web www.masterdatascienceucm.com
2. Enviar la documentación requerida a fin de evaluar la candidatura.
3. Entrevista con el solicitante.
4. Confirmación de selección.
5. Realización de un pago inicial.



Evaluación

Evaluación de candidaturas.



Pre-admisión

Confirmación como alumno del candidato.



Admisión

Confirmación de plaza y formalización de la matrícula.

Tanto la preinscripción como la pre matrícula quedan abiertas hasta comenzar el curso académico o completar plazas, estableciéndose lista de espera si procede. Los alumnos deberán ingresar 600 euros en concepto de pago inicial para el Máster Presencial y 500 euros en concepto de pago inicial para el Máster Semipresencial y el Máster Online, los cuales les serán descontados del importe total de la matrícula. En ningún caso se tendrá derecho a devolución de dicha cantidad, a excepción de que no se llegara a celebrar el curso.

Documentación requerida

Alumnos con titulación de **España**

Los documentos identificativos requeridos para la inscripción en el Máster son:

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

Alumnos con titulación de **Unión Europea**

- Currículum Vitae.
- Pasaporte/NIE (no válidas las cédulas de identificación de fuera de España).
- Título universitario (no es valido el certificado del título).
- Certificado oficial de notas.

*La documentación debe estar traducida al castellano por un traductor jurado homologado. (Solicitar listado oficial)

Alumnos con titulación de **Fuera de la Unión Europea**

- Currículum Vitae.
- Pasaporte/NIE (no válidas las cédulas de identificación de fuera de España).
- Título universitario legalizado con la Apostilla de la Haya (no es valido el certificado del título).
- Certificado oficial de notas.

*La documentación debe estar traducida al castellano por un traductor jurado homologado. (Solicitar listado oficial)



MODALIDADES

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





Máster Presencial



Duración

1 Año / 520 horas
60 ECTS

Inicio: Febrero de 2026
Fin: Febrero de 2027

Viernes: de 16:00 a 21:00 h
Sábados: de 09:00 a 14:00 h



Lugar

Facultad de Comercio y Turismo

Universidad Complutense de Madrid, Av. de Filipinas, 3, Chamberí, 28003 Madrid



Precio

6.700 €
+ 40 € de tasas de secretaría

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

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 de Formación Permanente de la UCM

Metodología Presencial

El curso se impartirá en aulas de la Universidad Complutense de Madrid, en la Facultad de Ciencias de la Documentación los viernes y sábados con masterclasses impartidas por diferentes expertos. La formación se realizará de forma tutorizada por los profesores. También 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.
- Vídeos de las clases y de casos prácticos.
- Tutorías online con el profesorado.
- Documentación, noticas y contenidos.
- Foro de los módulos del máster.
- Comunicación con los profesores vía mensajería.
- Chat entre alumnos.



Máster Semipresencial



Duración

1 Año / 520 horas
60 ECTS

Inicio: Octubre de 2026
Fin: Octubre de 2027

Jueves y viernes: de 16:00 a 21:00 h
Sábados: de 09:00 a 14:00 h



Lugar

Online con
presencialidad de 3
semanas

Facultad de Comercio y
Turismo

Universidad Complutense
de Madrid, Av. de Filipinas, 3,
Chamberí, 28003 Madrid



Precio

5.600 €
+ 40 € de tasas de
secretaría

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

*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 de Formación
Permanente de la UCM*

Metodología Semipresencial

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. La modalidad semipresencial contempla la realización de 3 semanas presenciales con masterclasses impartidas por diferentes expertos para preparar los TFM y hacer networking.

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.
- Vídeos de las clases y de casos prácticos.
- Tutorías online con el profesorado.
- Documentación, noticas y contenidos.
- Foro de los módulos del máster.
- Comunicación con los profesores vía mensajería.
- Chat entre alumnos.



Máster Online



Duración

1 Año / 520 horas
60 ECTS

Inicio: Febrero de 2026
Fin: Febrero de 2027



Lugar

Plataforma Online



Precio

4.650€
+ 40€ de tasas de secretaría

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

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 de Formación Permanente de la UCM

Metodología 100% Online

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.
- Vídeos de las clases y de casos prácticos.
- Tutorías online con el profesorado.
- Documentación, noticas y contenidos.
- Foro de los módulos del máster.
- Comunicación con los profesores vía mensajería.
- Chat entre alumnos.



Contacto

Teléfono de información

+34 687 30 04 04

Teléfono de admisiones

+34 667 89 05 83

Correo electrónico

info@masterdatascienceucm.com

Sitio Web

www.masterdatascienceucm.com

*La dirección del máster se reserva el derecho de modificar, suprimir y actualizar los profesores, la información y el programa del máster.





UNIVERSIDAD
COMPLUTENSE
MADRID

 nticmaster