



# Máster Oficial en Big Data Science

## Plan de Estudios

### MÓDULO 1: PROGRAMACIÓN Y COMPUTACIÓN

#### Python para análisis de datos (5 ECTS)

1. Sintaxis y estructuras de datos
2. Almacenamiento y manipulación de datos
3. Librerías Numpy, Pandas, Matplotlib y Seaborn
4. Proyectos

#### Bases de datos (2 ECTS)

1. Bases de datos relacionales
  - Modelo entidad relación
  - Normalización
  - SQL
2. Adquisición de datos
  - OLAP
  - Internet como fuente de datos
  - Intercambio de información
3. Almacenamiento distribuido
  - Blockchain
  - Hadoop (HDFS + MapReduce)
  - Procesamiento en tiempo real
4. Bases de datos NoSQL
  - Tipos
  - MongoDB
5. Google Cloud Platform
  - Compute
  - Cloud SQL
  - BigTable
  - DataStore
  - BigQuery

#### Visualización (2 ECTS)

1. Conceptos generales de visualización
2. Storytelling con datos
3. Plataformas comerciales para visualización

## **Técnicas de recogida de datos (2 ECTS)**

1. Data Management
  - Master Data Management (MDM)
  - Extracción de datos en entornos similares a empresa: SQL, Hive
2. Web scraping
3. Imágenes
4. Redes sociales

## **Técnicas de Big Data (3 ECTS)**

1. Arquitectura de Computadores. Cloud Computing. Infraestructura de Cloud. OCI
2. SQL analítico. Oracle Autonomous Database: ADW, ATP, JSON
3. Big Data Cloud Products
4. AWS

# **MÓDULO 2: ANÁLISIS DE DATOS**

## **Análisis estadístico de datos (8 ECTS)**

1. Revisión de probabilidad, variables aleatorias y contraste de hipótesis
2. Variables aleatorias multidimensionales. Densidad y masa conjunta. Distribuciones condicionales. Covarianza y correlación. Esperanza de un vector aleatorio, matriz de varianzas y covarianzas. Independencia de variables aleatorias
3. Análisis de la varianza
4. Regresión lineal múltiple y regresión logística
5. Regresión Lasso y Ridge
6. Análisis de componentes principales
7. Series temporales
8. Modelos de clasificación
9. Técnicas de agrupamiento

## **Preparación y limpieza de los datos (2 ECTS)**

1. Análisis exploratorio de datos
2. Pre procesado de datos
3. Detección de ruido y de outliers
4. Tratamiento de valores perdidos
5. Tratamiento del problema no balanceado
6. Datos estructurados y no estructurados
7. Evaluación de las distribuciones de las variables

## Machine Learning (6 ECTS)

1. Introducción a Machine Learning
2. Tipos de aprendizaje: supervisado, no supervisado, semi-supervisado y por refuerzo
3. Modelos frecuentistas vs bayesianos. Modelos paramétricos vs no-paramétricos
4. Inferencia vs predicción. Overfit vs underfit. Bias vs. variance
5. Tratamiento del dato, missing values e imputación. Feature engineering, feature importance y explicabilidad
6. Cadenas de Markov, Naive Bayes y reglas (sequence analysis y association analysis)
7. Instance based models (kNN), LDA, SVM, modelos basados en árboles (Decision Tree, Bagging Trees, Random Forest...) y regularización
8. Técnicas de clustering (kMeans, Jerárquico...) & Transformación de las dimensiones (Isomap, t-SNE, SOM, SVD, PCA...)
9. Network analysis: Spectral clustering, node centrality, bipartite network, co-citation, bibliographic coupling...
10. Análisis de supervivencia: censored and truncated data, Estimador Kaplan-Meier, log-rank test...
11. Ensemble learning methods (técnicas de ensemble secuenciales y en paralelo)
12. Casos prácticos: sistema de recomendación, series temporales y Datathon

## Deep Learning (3 ECTS)

1. Fundamentos de redes neuronales. Arquitecturas, funciones de activación y pérdida, capas y optimización de hyper-parámetros y modelos
2. Redes neuronales para datos tabulares: Clasificación, regresión y series temporales
3. Procesado de lenguaje natural. Clasificación de texto y clustering de documentos
4. Procesamiento de imagen
5. Transfer learning
6. Reinforcement learning
7. Procesado en la nube. Entrenamiento en paralelo de una red y desplegar modelos como servicio

A lo largo de la asignatura, desplegarás e integrarás los modelos de IA Generativa adaptados a tus aplicaciones y proyectos. Utilizarás todo su potencial para producir contenido personalizado y aprenderás el arte del *prompt engineering*. Asimismo, transformarás los LLMs con tus propios datos y conocerás las técnicas más avanzadas para crear prompts efectivos.

## MÓDULO 3: PROYECTOS

### Gestión de proyectos y visión de negocio (5 ECTS)

1. Planificación de proyectos: identificación, definición y objetivos.
2. Metodologías Agile
3. Privacidad, y transparencia. Ética de la inteligencia artificial
4. Proyectos de IA generativa
5. Aplicaciones

### Talleres con empresas (4 ECTS)

Exposición de ejemplos y casos de uso por parte de expertos de reconocidas empresas en varios sectores. Se abordan herramientas y técnicas impartidas durante el programa a través de proyectos reales y de actualidad.

## MÓDULO 4: TRABAJO FIN DE MÁSTER

### Trabajo Fin de Máster, TFM (18 ECTS)

El TFM consistirá en un trabajo original en el que deberán ponerse en práctica las competencias adquiridas durante el Máster. Se podrá hacer en grupos y desarrollar en el marco de una empresa o institución que proponga un proyecto de recogida, limpieza, preparación, analítica avanzada de datos y visualización de los resultados. También se podrá realizar a través de un proyecto de emprendimiento en este ámbito.

Se han de destacar aquellos aspectos éticos relativos al tratamiento de los datos, así como el impacto económico y social de los resultados. El alumno ha de demostrar que sabe planificar un proyecto y llevarlo a cabo en un entorno real de trabajo, de tal manera que adquiera una experiencia muy práctica en el ámbito de Data Science y Big Data.

