



Francisco Pérez Lozano, director académico del Instituto de Formación Continua de la Universidad de Barcelona,

### CERTIFICADO:

Que **Santiago Felipe Maggi Groetaers**, con número de DNI/Pasaporte P21204741, ha realizado el **Máster de Formación Permanente en Big Data & Data Science**, con una equivalencia de 60 créditos ECTS, durante el curso académico 2023/24 (impartido entre el 2 de noviembre de 2023 y el 17 de octubre de 2024), quedando a la espera de la entrega del título correspondiente.

Las calificaciones obtenidas son las siguientes:

Módulo	Créditos	Calificación
Herramientas de Big Data	9	9,13
Fundamentos de Estadística	5	9,50
Data Management& Data Digital	6	9,63
Técnicas avanzadas de Data Mining	5	9,83
Técnicas Avanzadas de Predicción	6	9,07
Técnicas de Machine Learning	6	9,60
Técnicas Avanzadas de Machine Learning	7	9,49
Project Management	6	9,51
Trabajo Final de Máster	10	9,30
Calificación final		9,42

- La calificación del curso está establecida en una escala de 0 a 10, siendo el 5 la calificación mínima para aprobar.
- El importe de la expedición del título está incluido en el precio de la matrícula.
- Curso propio diseñado según las directrices del Espacio Europeo de Educación Superior.
- 1 crédito ECTS equivale a 25 horas de trabajo del estudiante.

Y para que así conste, expido el presente certificado a petición de la persona interesada,



Barcelona, 31 de octubre de 2024



## Programa académico

### 1. Herramientas de Big Data

- 1.1. Almacenamiento de grandes volúmenes de datos: data mark, data lakes, definición de big data, SQL, NoSQL y procesamiento escalable: Hadoop, MapReduce y Spark
- 1.2. Herramientas en cloud: servicios en la nube. Diferencias entre las IaaS, PAAS, SaaS. Google Cloud, Azure Amazon Web Services
- 1.3. Monitorización y control. Control de versiones: Git o SVN. Ventajas e inconvenientes de utilizar GitHub. Alternativas a GitHub
- 1.4. Herramientas de análisis: programación en R y programación en Python
- 1.5. Herramientas de visualización: Power BI y Power Query (Microsoft Office Excel), Qlik y Tableau

### 2. Fundamentos de Estadística

- 2.1. Introducción a la Estadística: concepto de Estadística. conceptos básicos: población, muestra, variable estadística ...
- 2.2. Estadística Descriptiva Unidimensional: Distribución de frecuencias, métodos gráficos, tablas estadísticas. Medidas de posición, dispersión, forma y concentración. Gráficos.
- 2.3. Estadística Descriptiva Bidimensional: Distribución de frecuencias bidimensionales. Tablas estadísticas de doble entrada. Métodos gráficos.
- 2.4. Regresión y Correlación: Línea general de regresión. Correlación. Regresión no-lineal. Correlación entre atributos. Coeficiente Spearman, Kendall y Chi-cuadrado.
- 2.5. Introducción al Cálculo de Probabilidades: Azar y probabilidad. Definiciones y propiedades. Probabilidad condicionada e independencia. Teorema de Bayes y probabilidad total.

### 3. Data Management & Data Digital

- 3.1. Digital data: fuentes de datos externas oficiales (INE, EUROSTAT...)
- 3.2. Google Analytics. Social analytics: Hootsuite, Twitter analytics y Facebook insights
- 3.3. Web scrapping. Uso de correo basura
- 3.4. Data management. Gobierno de los datos. Ética y leyes
- 3.5. Data cooking. Preparación de los datos: selección, transformación y codificación
- 3.6. Data discovery: creación de nuevas métricas (KPI, insights), discretización de variables
- 3.7. Data quality. Técnicas avanzadas de detección de outliers, valores missing, imputación
- 3.8. Muestreo probabilístico, aleatorio, estratificado y sistemático

### 4. Técnicas avanzadas de Data Mining

- 4.1. Análisis multivariante: estimación, contraste de hipótesis e intervalos de confianza
- 4.2. Análisis de componentes principales (PCE) y factorial
- 4.3. Análisis discriminante: elección de variables (forward, backward, stepwise)
- 4.4. Análisis clúster: algoritmos de clasificación jerárquica y no jerárquica
- 4.5. Distribuciones. Gauss. Teorema Central de Límite

### 5. Técnicas Avanzadas de Predicción

- 5.1. Preparación de tablero de modelización. Modelos supervisados vs. modelos no supervisados
- 5.2. Técnicas para mejorar un modelo: training validation y cross validation
- 5.3. Regresión lineal y regresión logística
- 5.4. GLM (modelo lineal general). Ridge y PLS
- 5.5. Series temporales

### 6. Técnicas de Machine Learning

- 6.1. Classification algorithms: árboles de decisión
- 6.2. K-nearest neighbors
- 6.3. Model probabilistic: Bayes
- 6.4. Random forest
- 6.5. Boosting
- 6.6. Bagging
- 6.7. Adaboost

### 7. Técnicas Avanzadas de Machine Learning

- 7.1. Redes neuronales (neural networks NN)
- 7.2. Support vector machine (SVN)
- 7.3. Deep learning
- 7.4. Técnicas de machine learning en imágenes y video
- 7.5. Modelos de sentimientos (computación sobre texto, redes sociales, etc.)

### 8. Project Management





- 8.1. Herramientas para la gestión de clúster
  - 8.2. Validación del modelo: verificación p-values en regresiones; test de bondad de ajuste; R-cuadrado, MSE, BIC, AIC
  - 8.3. Test discriminate: curva ROC, curva lift, Gini index. Jackknifing y bootstrapping
  - 8.4. Diagnóstico de los residuos. Prueba de estabilidad y validación con prueba ciega
  - 8.5. Implementación del modelo: metodología para garantizar una solución viable
  - 8.6. Mantenimiento y actualización. Uso del modelo y requerimientos de implementación. Monitorizar el rendimiento y la efectividad de los modelos implementados
  - 8.7. Aportar análisis que faciliten la toma de decisiones y desarrollos futuros
  - 8.8. Gestión de proyectos: planificación de un proyecto
  - 8.9. Metodología agile y metodología scrum
- 9. Trabajo Final de Máster**

