

---

## GUÍA PARA REALIZAR EL TRABAJO FIN DE MASTER

---

---

**EL ALUMNO TENDRÁ QUE ELEGIR UNA DE LAS OPCIONES SIGUIENTES**  
**(nota: el trabajo es individual)**

---

### 1. ANÁLISIS DE UN DATASET (ORIENTACIÓN DATA SCIENTIST)

- a. Objetivos:
  - i. Analizar un dataset disponible públicamente (Kaggle, UCI Machine Learning Repository, Gapminder u otra fuente que el alumno considere siempre que el conjunto no tenga ningún derecho de uso).
- b. Fases Recomendadas:
  - i. El estudio y análisis de este dataset deberá de cumplir de forma general las fases de un proceso de modelización analítica estándar, entre las que se encuentran:
    - i. Crear un análisis descriptivo del conjunto (gráfico en lo posible).
    - ii. Realizar las transformaciones que se consideren más adecuadas o relevantes para el conjunto.
    - iii. Crear modelos de predicción utilizando diferentes técnicas de modelización (machine learning) justificando su uso, determinando el nivel de precisión y detallando las bondades, debilidades de cada técnica utilizada.
      - a. En este punto, se busca que el alumno proponga un desarrollo que aporte algo más que lo que se podría conseguir con un AutoML.
  - iv. Discusión de los resultados del modelo: explicatividad/interpretabilidad.
  - v. Realizar un informe final de conclusiones en el que las diferentes fases queden bien delimitadas y en particular donde las mejoras ofrecidas por el modelo queden bien explicitadas y las mejoras futuras que podrían plantear sobre el trabajo realizado.
    - a. Este informe final, tendrá una orientación tal que pueda ser entendida por un equipo de “Negocio”.
    - b. Podrá incluir elementos técnicos, pero deberá de incluir en mayor proporción detalles que expliquen los

resultados del modelo a una persona sin muchos conocimientos técnicos.

- vi. Además de las fases anteriormente descritas (propias de la metodología de modelización), se valorará muy positivamente, el que este modelo pueda productivizarse.
  - a. Entendemos por este aspecto el que el modelo pueda ser utilizado en un equivalente a una aplicación empresarial. Que al modelo se le puedan pasar nuevos valores y el modelo devuelva una predicción.

c. Extensión:

- i. La extensión total del trabajo no debe superar 20 caras (tamaño folio) sin contar los anexos, ni el índice de contenidos.
  - i. El tamaño de letra y el interlineado se deja a decisión del alumno, pero primando el sentido común y la legibilidad del documento (documentos a tamaño de letra 8 ó 9 o de 20 no tienen mucho sentido).
- ii. El código asociado y los estudios preliminares (la parte del descriptivo que considere el alumno) se aportarán como anexos. La extensión de estos anexos no cuentan para el tope de 20 caras comentado anteriormente.
- iii. El trabajo se puede realizar por entero en un notebook tipo Jupyter exportándose a formato HTML (caso de Jupyter)  
En estos casos:
  - i. por favor tened especial cuidado en no generar listados amplios de datos que no aportan valor.
- iv. Si el trabajo se realiza en el espacio de Colab, igualmente se ha de exportar el resultado a un .html para su correcta lectura.
  - i. En este caso, se puede adjuntar un link dentro del informe de conclusiones con la url utilizada de Colab.

d. Tecnologías:

- i. Lenguajes de programación Python.
  - i. Se valorará la legibilidad del código, el uso de comentarios y un correcto formateado.
- ii. Se recomienda el uso de un notebook: Jupyter

e. Visibilidad del trabajo si el conjunto es de Kaggle:

- i. Si el dataset elegido es de Kaggle, el código desarrollado se recomienda compartir como un “notebook” en el espacio asociados a los datos para este fin, incluyendo que el análisis forma parte de un proceso de evaluación del “Máster – XXXX”.

f. Realización de un video:

- i. Además de la documentación anteriormente descrita, se tendrá que realizar un video, en el que presentará su trabajo de una forma concisa destacando: conclusiones, lecciones aprendidas, etc.
- ii. El video tendrá una duración máxima de 5 minutos y se entregará en formato MP4.
- iii. En lo posible se espera que el video esté en un formato tal que no ocupe más de 50Mb.

**Atención:** Este video es propio del TFM. Si se solicitan otros videos, ya se indicará si se han de presentar varios o uno solo.

## 1. CREACIÓN DE UN PIPELINE DE PREPARACIÓN/DISPONIBILIZACIÓN DE DATOS (PERFIL DATA ENGINEER)

- a. Objetivos:
  - i. El objetivo consiste en preparar un pipeline, un conjunto de scripts que permitan realizar una/s ETLs (Extraction Transformation Loading) de diferentes fuentes e integrarlas en una base datos que pudiera ser utilizada para realizar un modelo.
  - ii. Estas ETLs deberán ser configurables en cuanto a la periodicidad de su ejecución y deberán contar con las soluciones necesarias para monitorizar su progreso/debugging.
- b. Tecnologías:
  - i. Cualquiera de las estudiadas en el Máster.
  - ii. Se puede optar por preparar el pipeline en una tecnología en particular o una combinación de Tecnologías.
- c. Documentación:
  - i. Se tendrá que documentar la arquitectura técnica elegida:
    - i. Sus componentes, sus inter-relaciones y las tecnologías empleadas en cada uno de estos elementos.
  - ii. Además de la solución técnica, la documentación deberá incluir detalles del caso de uso de negocio que solucionaría. Incluyendo referencias a alternativas existentes, diferenciando las mejoras que la propuesta introduce.
  - iii. En cuanto al código:
    - i. O bien se podrá incluir un repositorio GitHub o referir algún otro repositorio en la nube (Google, Amazon, Azure, etc).
  - iv. Además de la documentación, se deberá de incluir un video (o varios mini-videos) que expliquen la funcionalidad de cada uno de los módulos de la solución (demostrando cómo se realiza el proceso: captura de entradas y resultados de las salidas).
- d. Extensión:
  - i. En cuanto a la extensión de la solución, tampoco se espera que se presente una solución perfectamente disponible para un entorno empresarial, pero sí demostrar que la solución es perfectamente funcional de extremo a extremo.
  - ii. Que cumple el objetivo de la captura de diferentes fuentes de datos
  - iii. Y que éstos se disponibilizan en una/s tablas listas para ser explotadas: por procesos de modelización, de BI, etc.

## 2. EL ALUMNO PUEDE PROPONER UN TRABAJO QUE NO ENCAJE EN LAS PROPUESTAS ANTERIORES

- a. Objetivos:
  - i. El trabajo ha de estar relacionado alguno de los temas impartidos en el curso, pero siempre con una orientación de corte técnico.
  - ii. Que el trabajo implique el desarrollo de una solución software y que se pueda encuadrar en el ámbito de la analítica avanzada.
- b. Extensión:
  - i. La extensión total del trabajo no debe superar 20 caras (tamaño folio), con las mismas consideraciones comentadas en el punto 1 (también en el epígrafe de Extensión).
- c. Tecnologías:
  - i. Cualquiera de las impartidas en el Máster.
- d. Realización de un video:
  - i. Además de la documentación anteriormente descrita, se tendrá que realizar un video, en el que presentará su trabajo de una forma concisa destacando: enfoque, conclusiones, lecciones aprendidas, etc.
  - ii. El video tendrá una duración máxima de 5 minutos y se entregará en formato MP4.
  - iii. En lo posible se espera que el video esté en un formato tal que no ocupe más de 50Mb.

---

### Notas Generales:

- No se admitirán cambios de tema del TFM a menos de quince días para la fecha de entrega.
- El TFM se realizará ver detalles adjuntos (epígrafe de “Realización de los trabajos”).
- En el nombre del fichero se incluirá el nombre del alumno (Nombre y dos apellidos), separando el nombre y los apellidos con un guion bajo (“\_”):
  - Ejemplo: Maria\_Garcia\_Perez\_Estudio\_pajaros.zip
- Los tutores a cargo de mentorizar y corregir los trabajos serán **Carlos Ortega y Santiago Mota**.
  - Los tutores pueden ayudar en sugerir una orientación adecuada a una propuesta de trabajo, pero se evitará el enviar diferentes versiones del trabajo para confirmar si el enfoque o el nivel de avance, es el correcto.

### Realización de los trabajos:

- Los trabajos se realizarán en modalidad:

- Másters Online: Individual.

**Sobre el informe del TFM, a modo de resumen** la estructura del entregable sería:

- i. Documento (el de las 20 caras) que contiene:
  - i. el detalle del trabajo expuesto de una forma (a poder ser no muy técnica). Que incluye tablas resumen, uso de bullets para enumerar ideas, etc.
  - ii. En el texto se incluyen referencias a diferentes partes del Anexo donde se dan detalles más profundos de la idea expuesta.
- ii. Como Anexo se puede incluir:
  - i. El código desarrollado
  - ii. Estudio más detallados de por ejemplo el EDA, o de la ejecución de diferentes modelos.

Y además de este informe, en el entregable se incluirá el video.

---

#### **CHECKLIST:**

A modo de lista de comprobación de elementos importantes, se recomienda considerar lo siguiente:

- ¿Has mirado los derechos de uso de los datos?
- ¿Tienes el código compartido en un Github o en un Drive?
  - ¿Es accesible desde el link?
  - ¿Santiago Mota y Carlos Ortega tienen permisos de acceso?
- ¿La memoria ocupa 20 hojas?
- ¿Tienes el código en los Anexos?
- ¿El proyecto es reproducible?
- ¿Has incluido un apartado de conclusiones?
- ¿Has incluido el vídeo?
  - ¿Te has asegurado de que el video sea entre 5-6 minutos?.
  - ¿Te has asegurado de que el video describe tu proyecto (no es un elevator pitch)?.
- ¿Has incluido una breve lista (media cara) con la bibliografía y/o referencias?.