
UCM
**MÁSTER BIG DATA Y DATA SCIENCE. APLICACIONES AL COMERCIO, EMPRESA
Y FINANZAS 2019-2020 (CLASE 2)**

GUÍA PARA REALIZAR EL TRABAJO FIN DE MÁSTER

EL ALUMNO TENDRÁ QUE ELEGIR UNA DE LAS OPCIONES SIGUIENTES

1. ANÁLISIS DE UN DATASET

- 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:
 - 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.
 - iv. 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.

c. Extensión:

- i. La extensión total del trabajo no debe superar 20 caras (tamaño folio) sin contar los anexos.
- ii. El código asociado y los estudios preliminares se aportarán como anexos.
- iii. El trabajo se puede realizar por entero en un notebook tipo Jupyter o en un documento rMarkdown, exportándose en ambos casos a formato HTML (caso de Jupyter) o en formato HTML, pdf o Word para el caso de rMarkdown. En estos casos:
 - i. por favor tened especial cuidado en no generar listados amplios de datos que no aportan valor.

d. Tecnologías:

- i. Lenguajes de programación R / 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, fichero rMarkdown o una salida visual en Shiny.
- iii. También el informe se puede subir de forma gratuita a RPubs (acceso público que proporciona RStudio).

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

- i. Si el dataset elegido es de Kaggle, se valorará muy positivamente si el código/resultados se comparten como un “Kernel” 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 (opcional):

- i. Además de la documentación anteriormente descrita, se valorará muy positivamente, el que el alumno/grupo aporte un video en el que presentará su trabajo de una forma concisa (modo “Elevator-pitch”) destacando: conclusiones, lecciones aprendidas, etc.
- ii. El video tendrá una duración máxima de 5 minutos y se entregará en formato MP4.

2. ELEGIR UNA EMPRESA CONOCIDA

- a. Objetivos:
 - i. La empresa elegida debe ser una en la que el uso de tecnología BigData/Analítica Avanzada pueda mejorar significativamente sus resultados, su cuota de mercado, etc.
 - ii. O porque esta empresa no tenga desarrollada una estrategia de transformación digital y quiera modernizar sus procesos.
- b. Fases:
 - i. Justificar la elección de la empresa.
 - ii. Proponer el plan de transformación digital (alto nivel).
 - iii. Detallar la arquitectura BigData que se propone.
 - iv. Desarrollar el cambio organizativo asociado a la transformación: jerarquía, perfiles a desarrollar, relaciones entre los diferentes grupos, gestión del Gobierno del Dato, etc.
 - v. Detallar el conjunto de iniciativas a poner en marcha ordenadas de acuerdo con las necesidades del negocio.
- c. Extensión:
 - i. La extensión total del trabajo no debe superar 20 caras (tamaño folio).
 - ii. Se utilizará un tamaño de letra entre 10 y 11 puntos. Como interlineado se utilizará el que proporciona Word por defecto (1,5) u otro editor de textos equivalente.
 - iii. También los márgenes, superior, inferior y laterales serán los proporcionados por defecto por Word u otro editor de textos equivalente.
 - iv. En el trabajo se puede incluir un índice inicial de contenidos, aunque su extensión no se contabilizará en el número total de páginas.
 - v. Los gráficos y figuras que se incluyan deberán tener un pie de página que describa su contenido.
- d. Tecnologías:
 - i. Word, Google Docs.
- e. Realización de un video (opcional):
 - i. Además de la documentación anteriormente descrita, se valorará muy positivamente , el que el alumno/grupo aporte un video en el que presentará su trabajo de una forma concisa (modo “Elevator-pitch”) destacando: conclusiones, lecciones aprendidas, etc.
 - ii. El video tendrá una duración máxima de 5 minutos y se entregará en formato MP4.

Nota: Queda fuera del objetivo de esta opción, el presentar un trabajo en el que se proponga un plan de negocios de una nueva empresa (start-up), aunque se incluya una sección sobre su organización tecnológica.

3. 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.
- b. Extensión:
 - i. La extensión total del trabajo no debe superar 20 caras (tamaño folio).
- c. Tecnologías:
 - i. Word, Google Docs.
- d. Realización de un video (opcional):
 - i. Además de la documentación anteriormente descrita, se valorará muy positivamente , el que el alumno/grupo aporte un video en el que presentará su trabajo de una forma concisa (modo “Elevator-pitch”) destacando: conclusiones, lecciones aprendidas, etc.
 - ii. El video tendrá una duración máxima de 5 minutos y se entregará en formato MP4.

Notas Generales:

- El TFM se realizará en forma de grupo:
 - o los grupos serán de 6 alumnos.
- En el nombre del fichero se incluirá el nombre del grupo, separando con un guión bajo (“_”) el caso estudiado:
 - o Por ejemplo: Grupo_A_Estudio_mariposas.zip
- Los tutores o tutor se determinarán después de presentar la idea del tfm en la encuesta correspondiente antes del 1 de junio.
 - o 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.