



Máster en Data Science -- Normas para el Trabajo Fin de Máster

El Máster en Data Science requiere entregar un trabajo fin de máster que demuestre la capacidad del estudiante para realizar proyectos de data science, partiendo desde datos en bruto, y produciendo en el otro extremo una herramienta visual que sirva al usuario final para entender el resultado del análisis.

La memoria debe ser tan breve como sea posible. El objetivo principal de la memoria es servir de guía de replicación del trabajo. Es decir, otro data scientist, con el nivel de conocimientos adquirido en el máster, debería ser capaz de reproducir el trabajo usando la memoria como guía.

Se recomienda que la memoria incluya las siguientes secciones (pero no es obligatorio, y se puede estructurar de cualquier otra manera):

- Introducción. Descripción del problema que se va a resolver, de los datos que se van a usar, explicando por qué es un problema relevante. También se puede incluir una descripción de otros trabajos anteriores que el estudiante haya encontrado, y que estén relacionados con su TFM.
- Descripción de los datos de entrada. Formato. Diccionario de datos, describiendo los campos que contienen los datos, cómo están codificados, dónde se pueden encontrar. Si los datos son privados, y no se tiene permiso para publicarlos, hay que tener cuidado para no desvelar ninguna información sensible. No olvides que tu TFM estará disponible públicamente en Internet.
- Metodología. Esto incluye tanto una descripción genérica del proceso de datos que se ha realizado (sin excesivos detalles, para los detalles está el código que se escriba), como una descripción de cualquier método estadístico o de machine learning que se use
- Resumen del resultado. En esta sección se describiría qué se ha encontrado después del análisis
- Descripción/mini-manual de usuario del frontend que se haya desarrollado para ilustrar el resultado. Por ejemplo, si se ha creado un dashboard, debería incluirse unas instrucciones mínimas explicando cómo se usa el dashboard

Se puede emplear cualquier herramienta, lenguaje de programación, bibliotecas a las que los estudiantes tengan acceso. No es necesario usar ninguna herramienta o tecnología en concreto. Si se decide usar Spark, se recomienda usar el clúster disponible en el máster. En cualquier caso, la memoria debe aclarar qué tecnologías se están usando, qué paquetes se necesitan, qué dependencias hay, etc., de manera que se facilite la instalación de todas las herramientas necesarias, y sea posible replicar los resultados mostrados en la memoria.



Máster en Data Science -- Normas para el Trabajo Fin de Máster

Para la VIII edición, el trabajo fin de máster tiene que entregarse antes del 17 de enero de 2019 a las 23:59h. La forma de entrega es a través del repositorio de Github. Los profesores harán un pull de los repositorio en la fecha y hora límites. No se tendrán en cuenta commits posteriores.

Para elegir el tema del TFM, es necesario incluir un mensaje en Basecamp, con el asunto **[TFM] NOMBRE DE TU TFM**, incluyendo la URL del repositorio donde se desarrollará el TFM.

El trabajo lo pueden realizar una o dos personas. Si son dos personas, es necesario incluir el nombre de las dos personas en el mensaje en Basecamp. La calificación será la misma para las dos personas. La fecha límite para escribir el mensaje eligiendo tema y grupo es el 21 de noviembre de 2018.

No es necesario subir al repositorio los datos que se hayan usado. Si son datos privados, no olvidéis que el repositorio es público en Internet. Si los datos son públicos, se puede incluir un enlace o descripción de cómo obtener los datos. **No difundáis datos privados, sensibles, personales; en general, cualquier dato que no se pueda distribuir públicamente.**

Para la calificación del trabajo, se tendrán en cuenta los siguientes criterios:

- Claridad en la exposición de la memoria, de manera que sea posible replicar el trabajo fin de máster por un tercero con acceso a los mismos datos y herramientas
- Complejidad de los datos, del proceso de datos y del método de análisis aplicado
- Claridad, corrección y limpieza de todo el código que se escriba
- Relevancia y adecuación de los métodos de análisis y proceso usados para la cuestión planteada en la Introducción de la memoria
- Adecuación de las tecnologías elegidas al problema que se puede resolver
- Usabilidad y claridad de la visualización/frontend usado para comunicar los resultados. Se adoptará el punto de vista de un usuario final sin conocimientos técnicos, de Estadística, etc.

Los trabajos se puntuarán de 0 a 10. Si la entrega se retrasara, se descontarán 0.1 puntos por cada hora de retraso sobre la fecha y hora límite de entregas.