El material que habrá que entregar para la evaluación de la asignatura consiste en **una serie de funciones implementadas tanto en R como en Python**. Estas funciones están relacionadas con los diferentes ejercicios propuestos en los tutoriales. En todos los casos, **las funciones deberán estar adecuadamente documentadas**, tanto en lo que se refiere a la **información sobre las entradas y salidas** de la función como en lo referente al **código interno** de dicha función.

Además de la documentación de las funciones, la entrega deberá incluir un **documento** en el **que** se **ilustre el uso de las funciones**. En el caso de **R**, este documento deberá ser similar a **una viñeta**. En el caso de **Python**, el documento deberá ser **un notebook**. La entrega de estos documentos es un **requisito mínimo** para aprobar la asignatura, ya que será la principal forma de analizar la completitud y corrección del trabajo realizado. Por ese mismo motivo, es **importante** que el documento **refleje todas las funciones implementadas**. Además del ejemplo de uso (código), **se valorarán las explicaciones** sobre el uso de las funciones creadas.

El objetivo de la evaluación de la asignaturas es mostrar la **capacidad de programación** en R y Python, motivo por el cual el uso de implementaciones ya existentes se debe evitar al máximo. Es decir, **no está permitido usar paquetes o módulos** que implemente las funciones que se piden. Únicamente se podrán usar las funciones básicas de ambos lenguajes y paquetes o módulos para la visualización de datos y la gestión de estructuras de datos. En particular, las funcionalidades que se piden (discretización, entropía, información mutua, AUC, etc.) deben contar con implementaciones propias.

La funciones básicas a implementar son las indicadas a continuanción. Una adecuada implementación, junto con la documentación antes mencionada, permitirá alcanzar una nota máxima de 7. Más adelante se detallará como se pueden obtener los 3 puntos restantes. Funciones a implementar:

- Algoritmos de discretización para un solo atributo **y** para un dataset completo (ambas opciones): Igual frecuencia e igual anchura
- Cálculo de métricas para los atributos de un dataset: varianza y AUC para las variables contínuas y entropía para las discretas. La función deberá reconocer el tipo de atributo y actuar en consecuencia. Notese que en el caso del AUC, el dataset debe ser supervisado, es decir, es necesario especificar una variable clase binaria con la que evaluar el AUC de los atributos numéricos.
- Normalización y estandarización de variables, tanto de manera individual como para el dataset completo. Esto solo debe ser aplicado a atributos que sean numéricos.
- Filtrado de variables en base a las métricas implementadas. Es decir, partiendo de un dataset, obtener uno nuevo donde todas las variables cumplan los requisitos indicado (por ejemplo, una entropía superior a un cierto umbral).
- Cálculo de la correlación (información mutua en el caso de variables categóricas) por pares entre variables de un dataset. La función deberá considerar de que tipo es cada variable.
- Plots para el AUC y para las matrices de correlación/información mutua.

Más allá de la entrega básica, la nota se puede complementar con funcionalidades adicionales, ya sean en forma de nuevas funciones relacionadas o en forma de utilidades o mejoras. A modo de **ejemplo**, se considerarán funcionalidades adicionales:

- El diseño de objetos para gestionar los datasets y funciones de lectura y escritura. Es decir, la implementación de objetos, ya sea S3 o S4, para la gestión de los datasets.
- La creación de un paquete de R o un módulo de python instalable que facilite la distribución del software desarrollado. En estos casos es necesario que dicho software pueda ser instalado y cargado sin errores. En caso de crear un paquete de R, se valorará la inclusio del documento ilustrativo del uso del paquete como una viñeta directamente accesible desde la instalación.
- Creación de un repositorio en Github (o equivalente) en el que alojar el software. En el caso de R, si se ha creado un paquete se recomienda que este sea instalable directamente desde el repositorio usando el paquete **devtools**.
- Implementación de otros tipos de discretización.
- Funciones que incluyas más formas de visualización.
- Lectura y escritura de datasets en diferentes formatos y/o en un formato propio.
- Cualquier función que se considere de interés en el contexto de la gestión de datasets.