

---

***Todos los puntos del taller tienen el mismo peso***

1. Realice los siguientes ejercicios del libro “An Introduction to Statistical Learning with Applications in Python” (Link para descargar los conjuntos de datos: <https://www.statlearning.com/resources-python>):
  - Ejercicios Sección 4.8, página 193:
    - 5, 6, 7, 8, 12, 13, 16
  - Ejercicios Sección 5.4, página 224:
    - 3, 5, 8
  - Ejercicios Sección 12.6, página 552:
    - 3, 10, 13
2. Considere el conjunto de datos de fraudes realizados con tarjeta de crédito en el siguiente [link](#). Utilizando los métodos vistos en clase, encuentre el mejor modelo de clasificación para el caso en estudio, si una transacción es fraudulenta o no. Considere:
  - Eliminación de información redundante
  - Selección de variables
  - Selección de una métrica adecuada
  - Definición de métricas de un modelo baseline (clasificador aleatorio) como punto de referencia
  - Entrenamiento, validación y prueba para los modelos
  - Ajuste de hiperparámetros para cada uno de los modelos
  - Selección final del modelo y su desempeño esperado en producción
3. Considere el conjunto de datos ‘*data\_gapminder\_experiment.csv*’, en el que se encuentra información socio-demográfica de diferentes países para el año 2015. Aplique una estrategia de métodos jerárquicos, K means y DBScan para agrupación de países. Analice los resultados y determine qué estrategia es la más adecuada para el problema en estudio. Considere:
  - Interpretación/Análisis de clusters. ¿Qué podrían significar los clusters?
  - Grupos pertinentes a encontrar
  - Representación visual de los grupos utilizando estrategias reducción de dimensionalidad. ¿Al utilizar componentes principales, qué interpretación pueden tener los primeros dos componentes?
  - Métrica adecuada de validación de la composición de los clusters
  - Análisis de los resultados

**Entregable:** Se recibirá un trabajo por equipo (equipos creados en EAFIT Interactiva), en el cual debe de contener (1) informe/documento escrito que contenga el enunciado de cada punto, además de su respectiva solución y análisis, adicionalmente (2) el código organizado y documentado utilizado para dar solución a la implementación computacional realizada en el taller. Código que no ejecute o compile, se tomará como razón para disminuir la puntuación de los ejercicios computacionales.

**Fecha de entrega:** 29 de noviembre 2024.