

**Descripción General:** el proyecto consiste en resolver un problema sobre un conjunto de datos. Se debe analizar el contexto, hacer un análisis exploratorio de datos, y seleccionar el mejor modelo al evaluar diferentes técnicas de regresión lineal, polinómica a trazos, y splines.

**Directrices:** el proyecto consistirá en

1. Estudiar el contexto del problema a resolver.
2. Una primera parte del proyecto, consiste en realizar un análisis exploratorio de los datos, en donde se realicen conclusiones, mediante gráficas y cálculos estadísticos básicos.
3. Realizar una informe en el que se describan claramente el problema, el desarrollo, análisis, metodología y procedimientos desarrollados y las pertinentes conclusiones que obtiene en el contexto del problema considerado.
4. Realizar una presentación final en la que se describirá brevemente el problema, los métodos utilizados, los resultados obtenidos, el análisis realizado, y las conclusiones pertinentes.

**Requerimientos:** el proyecto debe:

1. Desarrollarse en Python.
2. Realizar una análisis exploratorio de datos.
3. Realizar modelos de regresión múltiple (lineal, Ridge, Lasso) en donde se respondan las siguientes preguntas, adaptadas al contexto de problema:
  - a) ¿Al menos uno de los predictores (descriptores) es útil en predecir la variable respuesta?
  - b) ¿Todos los predictores ayudan a explicar la variable respuesta, o tan sólo un conjunto de predictores es útil?
  - c) ¿Qué tan bien se ajusta el modelo a los datos?
  - d) Dado un conjunto de valores para los predictores, ¿Qué valor de la variable respuesta se predice y qué tan exacta es la predicción?
4. Evaluar y seleccionar los modelos usando validación cruzada. Y posteriormente, evaluar el modelo sobre un conjunto independiente de prueba.
5. Realizar modelos de ajuste con base en polinomios y splines.
6. Realizar estimaciones de intervalos de confianza usando bootstrap.
7. Concluir con las observaciones generadas a través del análisis, creación y evaluación de los modelos construidos.
8. El trabajo deber estar bien escrito, sin errores gramaticales y ortográficos.
9. El trabajo debe ser escrito en LATEX (entregar fuentes) y debe estar acompañado adicionalmente de un notebook (jupyter notebook).

**Entregables y Fechas:**

---

**Entrega:** Documento de **a lo sumo cinco páginas** con la descripción del problema considerado, el análisis exploratorio de datos, el conjunto de técnicas utilizadas y los procedimientos desarrollados, una discusión y análisis de los resultados y las conclusiones obtenidas. El documento deberá incluir un **resumen ejecutivo** de a lo sumo una página donde resuma brevemente el problema considerado y las conclusiones.

**Fecha de entrega:** a más tardar el **24 de agosto de 2021**.

**Puntaje:** 70 puntos

Informe ejecutivo, descripción del problema, análisis exploratorio y calidad de las conclusiones: 20 puntos

Descripción del procedimiento, análisis y discusión realizados: 30 puntos.

Claridad argumentativa, ortografía, presentación: 10 puntos.

Código en python **en notebook y documentado** que respalde todos los análisis/procedimientos y permita replicar todos los resultados: 10 puntos

**Presentación/Defensa:** video de presentación donde se describa el problema, el conjunto de técnicas seleccionadas y procedimientos desarrollados, el análisis de los resultados y las conclusiones obtenidas. Se abrirá un Foro en E-aulas donde se realizará la defensa mediante preguntas escritas u orales.

**Fecha:** **24 de agosto de 2021**.

**Puntaje:** 30 puntos

#### **Otros:**

**Grupos:** el proyecto se realizará en grupos de **2 o 3 personas**. En ningún caso un grupo podrá estar conformado por más de 3 personas ni se aceptarán trabajos individuales.

**Caso especial:** si se detecta que un estudiante no participó en la realización de su proyecto, su nota será cero.