

1. Ejercicio Práctico 1: Clasificación de Hábitos de Fumadores

Contexto: Te encuentras como analista de datos en el laboratorio "HealthData Lab", un referente en investigación epidemiológica y análisis de datos de salud pública. Tu equipo ha recopilado datos de salud de individuos en un estudio observacional con el fin de identificar patrones relacionados con el hábito de fumar.

El desafío: Utilizar este conjunto de datos para clasificar a los individuos según sean fumadores o no fumadores. Esta clasificación ayudará en el desarrollo de campañas de salud pública y en la comprensión de los impactos del tabaquismo en la salud general.

Objetivo: Basándote en el conjunto de datos proporcionado, tu tarea es clasificar cada registro como fumador o no fumador. La categorización deberá estar fundamentada en un análisis exploratorio de los datos y las características observadas.

2. Instrucciones:

1. Exploración Inicial de Datos (1 punto):

- Revisar el conjunto de datos para comprender su estructura y contenido.
- Identificar tipos de variables y realizar estadísticas descriptivas básicas.
- Visualizar la distribución del hábito de fumar en función del género mediante histogramas.

2. Feature Engineering (2 puntos):

- Ejecutar técnicas de binarización y agrupación por cuantiles en las variables continuas.
- Crear nuevas columnas en el Dataframe basadas en las transformaciones realizadas.

3. Preprocesamiento de Datos (1 punto):

- Tratar valores perdidos y atípicos.
- Dividir el conjunto de datos en entrenamiento y prueba.

4. Análisis Exploratorio de Datos (EDA) (1 punto):

- Profundizar en el análisis de las variables y su relación con la variable objetivo.
- Estudiar las correlaciones para identificar posibles predictores.

5. Selección de Características (1 punto):

- Utilizar métodos estadísticos o algoritmos para identificar las características más relevantes.

6. Construcción del Modelo (1 punto):

- Elegir entre SVM o kNN para el desarrollo del modelo
- Realizar el entrenamiento con el conjunto de datos de entrenamiento.

7. Evaluación del Modelo (2 puntos):

- Medir el rendimiento del modelo con métricas apropiadas.

- Aplicar validación cruzada para la optimización de hiperparámetros.

8. Interpretación de Resultados y Conclusiones (1 punto):

- Analizar la importancia de las características y el comportamiento del modelo.
- Proponer mejoras y reflexionar sobre los resultados.

3. Entrega:

- Se espera que completes el NoteBook adjunto al presente documento (clasificación_fumadores.ipynb) abarcando todas las fases mencionadas anteriormente. El informe (Notebook) debe incluir el código en Python empleado, gráficos, resultados y conclusiones.
- Solo debe ser entregado el documento 'clasificación_fumadores.ipynb' con la extensión de nombre y apellidos. Ejemplo: clasificación_fumadores_Juan_Cuesta.ipynb

4. Instrucciones y anotaciones extra:

- En el notebook proporcionado (clasificación_fumadores.ipynb), se detallan de manera precisa las instrucciones y requerimientos para cada ejercicio.
- Es fundamental que todas las respuestas, comentarios y análisis, especialmente aquellos relacionados con los gráficos, estén adecuadamente justificados.
- Esta práctica es de carácter **individual**, por lo que no se permite el trabajo en grupo.
- La fecha límite para la entrega de la práctica es el **20 de febrero**, incluido este día.