

Laboratorio 10.
Análisis Exploratorio de Datos y Modelado con ChatGPT

INSTRUCCIONES:

Objetivos:

- Familiarizarse con el Análisis Exploratorio de Datos (EDA) utilizando Python.
- Aprender a desarrollar y evaluar modelos de aprendizaje automático.
- Reflexionar sobre la aplicabilidad de estas habilidades en contextos profesionales.

Herramientas:

- Python
- Bibliotecas: Pandas, Matplotlib, Seaborn, Scikit-learn
- ChatGPT para consultas y asistencia en codificación.

Conjunto de Datos:

Utilizaremos datos de la "Encuesta Nacional de Condiciones de Vida 2014" (ENCOVI) de Guatemala. Estos datos deben ser limpiados y preprocesados antes del análisis.

Parte 1: Análisis Exploratorio de Datos (40 Puntos)

1. Carga y Limpieza de Datos (10 puntos):

- Importar el conjunto de datos.
- Identificar y tratar valores faltantes.
- Normalizar nombres de columnas.

2. Análisis Univariable (10 puntos):

- Estadísticas descriptivas (media, mediana, moda).
- Distribuciones de variables importantes mediante histogramas o boxplots.

3. Análisis Bivariable (10 puntos):

- Correlaciones entre variables.
- Scatter plots para relaciones entre pares de variables.

4. Visualización de Datos (10 puntos):

- Utilizar gráficos de barras, pie y heatmaps para representar relaciones y distribuciones.

Parte 2: Modelado y Pruebas (45 Puntos)

1. Preparación de Datos (5 puntos):

- Dividir el conjunto de datos en entrenamiento y prueba.
- Estandarización o normalización de variables si es necesario.

2. Construcción del Modelo (20 puntos):

- Elegir un modelo de aprendizaje automático adecuado para la tarea.
- Justificar la elección del modelo.
- Entrenar el modelo con el conjunto de datos.

3. Evaluación del Modelo (10 puntos):

- Utilizar métricas adecuadas para evaluar el modelo (e.g., precisión, recall, F1-score).

4. Optimización y Ajuste (10 puntos):

- Ajustar hiperparámetros.
- Evaluar la mejora en el desempeño del modelo.

Parte 3: Reflexión (5 Puntos)

- Redactar una reflexión sobre la experiencia en el laboratorio, incluyendo:
 - Desafíos encontrados y cómo se superaron.

- Lecciones aprendidas.
- Potencial de aplicación de estas habilidades en el ámbito profesional como Científico de Datos.

Rúbrica de Evaluación (100 puntos en total)

- **Carga y Limpieza de Datos: 10 puntos**
- **Análisis Univariable: 10 puntos**
- **Análisis Bivariable: 10 puntos**
- **Visualización de Datos: 10 puntos**
- **Preparación de Datos para Modelado: 5 puntos**
- **Construcción del Modelo: 20 puntos**
- **Evaluación del Modelo: 10 puntos**
- **Optimización y Ajuste del Modelo: 10 puntos**
- **Reflexión y Aplicabilidad: 5 puntos**
- **Entrega y Formato de Reporte: 10 puntos** (incluye claridad, ortografía, y formato adecuado del reporte y código).

Entrega Final:

La entrega final debe incluir:

- Un reporte escrito explicando cada paso realizado, incluyendo código y visualizaciones.
- Un notebook de Jupyter con el código ejecutable y comentarios explicativos.
- La reflexión sobre la experiencia del laboratorio.