

Actividad Practica

Stephanie Ríos Rodríguez

Fundación Universitaria Compensar  
Programación para Ciencia de Datos II

Docente

Sebastián Rodríguez Muñoz

Bogotá D.C, Colombia

28 de marzo de 2024

## Actividad Practica

Stephanie Ríos Rodríguez

Docente

Sebastián Rodríguez Muñoz

## Actividad # 6

Fundación Universitaria Compensar

Facultad De Ingeniería-Ciencia De Datos

Programación para Ciencia de Datos II

Bogotá D.C, Colombia

28 de marzo de 2024

## Contenido

Introducción.....	4
Desarrollo.....	5
Conclusiones.....	9

## **Introducción.**

Este trabajo se centra en el análisis de los datos de colombianos detenidos en el exterior, con el fin de comprender si existen diferencias relevantes entre los registros de Chile y Estados Unidos. La finalidad principal es transformar información compleja en visualizaciones intuitivas mediante un dashboard interactivo. Con este dashboard, se puede explorar la cantidad promedio de detenidos en cada país y observar cómo varían estos datos según la selección realizada. De esta manera, el proyecto busca no solo revelar diferencias en los patrones de detención, sino también ofrecer una herramienta práctica para apoyar la toma de decisiones en contextos relacionados con la gestión de datos y políticas de seguridad.

### **Desarrollo.**

El propósito de esta actividad es desarrollar sus habilidades para presentar los resultados que se han obtenido a lo largo del curso. Con esto en mente el entregable que se estará evaluando es un dashboard con los principales descubrimientos sobre la problemática que fue planeada al inicio. Teniendo esto en cuenta, se realizan las siguientes recomendaciones para mejorar el modelo que se ha estado trabajando.

1. Evalúe los resultados que ha obtenido en relación con la problemática que se definió durante la primera actividad. Asegúrese, que cada uno de los análisis que haya planteado sea con el objetivo de mejorar la comprensión que se tiene del problema.
2. Realice una evaluación del rendimiento que presenta el modelo actual, la habilidad que tiene de ofrecer información sobre la problemática. Defina si debe mejorar el modelo iterativamente, o se debe replantear el enfoque de la pregunta inicial para obtener mejores resultados.
3. Considere la recopilación de datos adicionales o un procesamiento adicional de acuerdo a las técnicas que se vieron en el apartado teórico de esta etapa. Puede realizar diferentes y comprobar como mejora el rendimiento dependiendo de la calidad de los datos utilizados.
4. Experimente con los hiperparámetros que definen el modelo. Es importante que tenga métricas que le permitan evaluar el desempeño de los mismos con el objetivo de optimizar los resultados obtenidos. experimentación controlada con los hiperparámetros es una práctica recomendable.
5. Considere si alguno de las métricas del modelo mejoraría información del algún aspecto del problema. La mejora de la capacidad predictiva se puede lograr mediante la creación de nuevas características o la adaptación de las existentes.
6. Considere la Regularización: En caso de que se detecte un sobreajuste en el modelo, se debe considerar la aplicación de técnicas de regularización. Estas implican la incorporación de términos de penalización en la función de pérdida.

## **Resultados Obtenidos y Contribuciones al Entendimiento del Problema:**

- **Contraste de Hipótesis:**

- Se formuló la hipótesis nula ( $H_0$ : las medias son iguales) y la alternativa ( $H_1$ : las medias difieren) para comparar la cantidad de detenidos entre países.
- Los resultados (estadístico t y p-valor) indicaron diferencias significativas, lo que respalda la idea de que el país de detención influye en la cantidad de detenidos.

- **Análisis de Regresión:**

- La regresión lineal aplicada entre la fecha (convertida a un valor ordinal) y la cantidad de detenidos mostró una pendiente casi nula y un  $R^2$  muy bajo, lo que evidencia que la fecha por sí sola no explica la variabilidad de los detenidos.
- En la práctica, se observó que, aunque la mayoría de los registros son bajos (mediana de 1), existen outliers o casos extremos que elevan la media y la desviación estándar, sugiriendo una distribución asimétrica.

- **Modelo de Regresión Logística:**

- Al convertir la variable “CANTIDAD” en una clasificación binaria (alta vs. baja cantidad), se evidenció que el modelo, utilizando únicamente el valor ordinal de la fecha, no fue capaz de detectar la clase positiva, lo que apunta a la necesidad de incluir más variables o mejorar el tratamiento de los datos.

## **Fortalezas y Limitaciones del Modelo Actual:**

- **Fortalezas:**

- El contraste de hipótesis confirma, con evidencia estadística, la influencia del país en la cantidad de detenidos, lo cual es un hallazgo relevante que respalda la hipótesis inicial.
- Los análisis descriptivos y visualizaciones (histograma, gráfico de dispersión con línea de regresión) permiten identificar tendencias generales y detectar outliers.
- **Limitaciones:**
  - El modelo de regresión lineal no logra explicar la variabilidad de la cantidad de detenidos, dado que utiliza pocas variables (por ejemplo, solo la fecha).
  - La regresión logística, al usar un predictor único, clasifica prácticamente todos los casos en la clase negativa, lo que indica un problema de desbalanceo o insuficiencia en la información predictiva.
  - Es probable que se requiera incorporar variables adicionales (como tipo de delito, ubicación geográfica detallada o características demográficas) o aplicar transformaciones a las actuales para capturar de forma más completa la complejidad del fenómeno.

### **Conclusión para el Punto 1:**

- **Evaluación Integral:**

Los análisis realizados hasta ahora han permitido confirmar que existen diferencias significativas en la cantidad de detenidos en función del país de detención y han mostrado que, con la información y variables actuales, el modelo predictivo (tanto en regresión lineal como logística) presenta limitaciones importantes para explicar la variabilidad del fenómeno.
- **Relación con la Problemática Inicial:**

Estos hallazgos indican que, para entender de manera profunda la situación de los detenidos, ya sea en la comparación entre países o en el análisis de los colombianos detenidos en el exterior, es necesario ampliar la cantidad de variables y posiblemente mejorar el preprocesamiento de los datos. Esto, a su vez, abrirá la puerta a optimizar y/o

replantear el enfoque inicial, lo cual será esencial para desarrollar un dashboard que no solo muestre resultados, sino que también permita tomar decisiones informadas.



## **Conclusiones.**

- ✓ El análisis muestra que existen diferencias significativas en la cantidad promedio de detenidos entre Chile y Estados Unidos. Estos hallazgos sugieren que el país de detención influye en los registros y pueden estar asociados a diferentes políticas o procedimientos.
- ✓ El dashboard interactivo permite filtrar la información por país, facilitando la comparación directa entre los datos de ambos lugares. Esto demuestra que la visualización dinámica es una herramienta eficaz para comprender y analizar grandes volúmenes de datos de manera sencilla.
- ✓ Aunque se logró identificar diferencias en los promedios, el análisis indica que utilizar únicamente las variables actuales no permite captar la complejidad completa del fenómeno. Es posible que factores adicionales, como el tipo de delito o características demográficas, puedan aportar una comprensión más profunda.
- ✓ Los resultados obtenidos abren la puerta a nuevas líneas de investigación. Se plantea que, al incorporar más variables y aplicar técnicas de análisis más avanzadas, se podría mejorar la capacidad de explicar las variaciones en la cantidad de detenidos. Esto permitiría desarrollar modelos predictivos más precisos y apoyar de manera más efectiva la toma de decisiones en el ámbito de la seguridad y la política migratoria.