
1. INDICACIONES

- En esta actividad se evalúa si el estudiante (*Criterio 3.1: Aplica modelos de aprendizaje no supervisado en casos prácticos complejos, analizando los resultados y proponiendo mejoras basadas en métricas de rendimiento.*

2. DESCRIPCIÓN

En el contexto de un aumento significativo de la violencia en el país, la Policía Científica busca mejorar sus estrategias de prevención y acción a través del análisis de datos. Como analistas de datos, tu misión es identificar patrones significativos y reducir la complejidad de los datos para extraer información valiosa.

Pregunta esencial

- ¿Cómo pueden las técnicas de clustering y reducción de dimensiones ayudar a identificar patrones ocultos en los datos y facilitar decisiones informadas en el manejo de eventos violentos?

Reto

Como parte de un equipo de análisis de datos, tu misión es desentrañar patrones clave dentro de un conjunto de datos complejos sobre eventos violentos en el país. El reto consiste en aplicar técnicas de clustering y reducción de dimensiones para extraer información valiosa y visualmente clara que pueda ser utilizada por la Policía Científica para mejorar sus estrategias. Tu objetivo final será entregar:

- Un Jupyter Notebook replicable, que incluya código y visualizaciones técnicas para explorar y analizar los datos.
- Un informe divulgativo para la Policía Científica que presente los hallazgos clave de manera clara y accesible, con visualizaciones comprensibles.

Preguntas guía

- ¿Cómo se preprocesan los datos para su análisis efectivo?
- ¿Qué algoritmos de clustering son más adecuados para este tipo de datos?
- ¿Qué criterios usarás para evaluar la calidad de los clusters obtenidos?
- ¿Cómo puede la reducción de dimensiones facilitar la visualización de patrones complejos?
- ¿Cómo puedes comunicar los hallazgos de manera efectiva a un público no técnico?

Actividades guía

1. Investigar las técnicas de clustering (como K-Means, DBSCAN) y reducción de dimensiones (PCA, t-SNE).
2. Preprocesar el dataset:
 - Manejo de valores nulos.
 - Estandarización de variables.
 - Codificación de variables categóricas.
3. Aplicar clustering para identificar agrupaciones significativas en los datos.
4. Evaluar los agrupamientos para obtener el mejor posible.
5. Implementar reducción de dimensiones para visualizar las relaciones entre variables y clusters.
6. Generar visualizaciones técnicas y explicativas:
 - Gráficos de dispersión y mapas de calor.
 - Representaciones tridimensionales para datos complejos.
7. Redactar un informe final, estructurando los hallazgos en:
 - Introducción al problema.
 - Metodología aplicada.
 - Resultados obtenidos y su interpretación.
 - Recomendaciones basadas en los hallazgos.

Recursos

- Dataset confidencial proporcionado por la Policía Científica (previo acuerdo de confidencialidad).
- Documentación de Scikit-learn, Pandas y Matplotlib.
- Tutoriales y ejemplos sobre clustering y reducción de dimensiones dados en clases.

PRODUCTO

2.1 Jupyter Notebook

El *Jupyter Notebook* debe ser técnico, bien documentado y fácilmente replicable, permitiendo que otros analistas reproduzcan el análisis con datasets similares. Debe estar generado en el [formato base](#) dado en clases.

- 1. Título y descripción inicial:** Incluye un título descriptivo, los autores y una breve introducción que explique el problema, los objetivos del análisis y las técnicas utilizadas.

- 2. Carga y descripción de datos:** Código para cargar el dataset proporcionado. Análisis exploratorio inicial que describa las variables más relevantes y sus características (tipos, distribuciones, etc.).
- 3. Preprocesamiento de datos:** Limpieza del dataset, incluyendo manejo de valores nulos, codificación de variables categóricas y normalización. Explicación de los pasos tomados y justificación de las decisiones.
- 4. Análisis con clustering:** Aplicación de algoritmos de clustering, como K-Means, DBSCAN o jerárquico. Visualizaciones técnicas que representen los resultados, como gráficos de dispersión coloreados por cluster.
- 5. Reducción de dimensiones:** Implementación de técnicas como PCA o t-SNE para reducir las dimensiones de los datos. Visualización en 2D o 3D de los datos en el espacio reducido, resaltando los patrones principales.
- 6. Interpretación de resultados:** Discusión breve de los clusters formados y cómo las técnicas aplicadas ayudan a entender los datos.
- 7. Conclusión técnica:** Resumen de los hallazgos técnicos y limitaciones del análisis.

2.2 Informe Divulgativo

El informe divulgativo debe estar dirigido a un público no técnico, como responsables de la toma de decisiones en la Policía Científica. Debe estar hecho en \LaTeX , en el [formato proporcionado](#), e incluir bibliografía relevante (tanto que vincule la aplicación de las técnicas con el tipo de datos).

- 1. Introducción:** Contexto del problema, relevancia del análisis y objetivos de la investigación.
- 2. Metodología:** Breve descripción de las técnicas utilizadas (clustering, reducción de dimensiones) y su utilidad.
- 3. Resultados principales:** Explicación de los clusters encontrados y sus características principales. Gráficos comprensibles que muestren patrones relevantes, como la distribución de eventos por tipo o área.
- 4. Implicaciones prácticas:** Interpretación de los resultados en términos de acciones concretas para la Policía Científica.
- 5. Recomendaciones:** Sugerencias basadas en el análisis, como estrategias de intervención o necesidades de datos adicionales.
- 6. Conclusión:** Resumen breve de los hallazgos y su importancia para la resolución del problema.

3. RÚBRICA DE EVALUACIÓN

Jupyter Notebook (30 puntos totales)

1. Título y descripción inicial (2 puntos)

- Introducción que contextualiza el problema y los objetivos (1 punto).
- Breve descripción de las técnicas utilizadas (1 punto).

2. Carga y descripción de datos (4 puntos)

- Descripción de las variables relevantes, análisis exploratorio básico (3 puntos).
- Identificación y representación gráfica adecuada de distribuciones (1 punto).

3. Preprocesamiento de datos (5 puntos)

- Limpieza de datos correctamente implementada (3 puntos).
- Justificación clara de las decisiones tomadas en el preprocesamiento (2 puntos).

4. Análisis con clustering (8 puntos)

- Implementación adecuada de al menos dos algoritmos de clustering (4 puntos).
- Visualizaciones técnicas que representen los resultados (2 puntos).
- Explicación de la calidad de los clusters obtenidos (2 puntos).

5. Reducción de dimensiones (4 puntos)

- Implementación de una técnica de reducción de dimensiones (2 puntos).
- Visualizaciones técnicas en 2D o 3D que resalten patrones clave (2 puntos).

6. Interpretación de resultados (2 puntos)

- Discusión clara de los clusters formados y los patrones identificados (1.5 puntos).
- Relación de los resultados con el problema inicial (0.5 puntos).

7. Conclusión técnica (2 puntos)

- Resumen conciso de los hallazgos técnicos (1.5 puntos).
- Identificación de limitaciones y posibles mejoras (0.5 puntos).

8. Legibilidad y simpleza del código (3 puntos)

- Código bien comentado, organizado en secciones claras (1.5 puntos).
- Uso de celdas Markdown para explicar los pasos (1.5 puntos).

Informe Divulgativo (20 puntos totales)**1. Introducción (3 puntos)**

- Contextualiza el problema y la relevancia del análisis (2 puntos).
- Explica claramente los objetivos de la investigación (1 punto).

2. Metodología (2 puntos)

- Breve descripción de las técnicas utilizadas (1.5 puntos).
- Explicación de cómo estas técnicas ayudan a entender los datos (0.5 puntos).

3. Resultados principales (5 puntos)

- Explicación de los clusters encontrados y sus características (2.5 puntos).
- Uso de gráficos comprensibles y relevantes para el público objetivo (2.5 puntos).

4. Implicaciones prácticas (3 puntos)

- Relación de los resultados con acciones concretas (2 puntos).
- Interpretación de patrones significativos y su importancia (1 punto).

5. Recomendaciones (2 puntos)

- Propuestas claras y aplicables basadas en los resultados del análisis (2 puntos).

6. Conclusión (3 puntos)

- Resumen breve y relevante de los hallazgos (2 puntos).
- Importancia de los resultados para la resolución del problema (1 punto).

7. Conclusión (3 puntos)

- Diseño profesional y organizado con secciones bien definidas (1.5 puntos).
- Uso de gráficos y tablas adecuadamente formateados (1.5 puntos).