|  |
| --- |
| **Nombre del proyecto** |
| **Segmentación de Usuarios - Aprendizaje No Supervisado** |

|  |  |
| --- | --- |
| **Profesor** | **Gladys María Villegas Rugel** |
| **Materia** | **Aprendizaje Automático** |
| **Alumno** | **- MARÍA AUGUSTA FLORES**  **- MARCELO ISMAEL ANDRADE**  **- JORGE ANDRÉS ORELLANA**  **- GALO VLADIMIR GONZÁLEZ** |
| **Fecha** | **Mayo 12 del 2025** |

|  |
| --- |
| **Resumen del Proyecto** |
| **Introducción**  Las plataformas digitales recopilan grandes volúmenes de datos de sus usuarios. Este proyecto busca aprovechar esos datos para identificar perfiles de comportamiento utilizando algoritmos de clustering, sin necesidad de etiquetas previas. A través de este enfoque, se pueden generar estrategias personalizadas para grupos de usuarios similares.  Este proyecto tiene como objetivo segmentar perfiles de usuarios utilizando modelos de aprendizaje no supervisado. Se analizaron datos de comportamiento de clientes de una tienda online, aplicando técnicas como K-means, DBSCAN y PCA para descubrir agrupamientos útiles para la toma de decisiones comerciales.  **Dataset Utilizado**   * **Nombre:** Online Retail Dataset (UCI Machine Learning Repository) * **Observaciones:** : transacciones de clientes en una tienda online del Reino Unido entre 2010 y 2011. * **Formato:** convertido de Excel a CSV para facilitar la carga en Python. * **Ubicación del dataset:** Carpeta data/online\_retail\_norm.csv   **Objetivos de la Actividad**   * Aplicar modelos no supervisados (K-means, DBSCAN) para identificar grupos de usuarios. * Utilizar PCA para visualizar agrupamientos en dos dimensiones. * Generar una tabla resumen que describa los clusters detectados. * Documentar el proceso completo con notebooks reproducibles.  **Metodología Aplicada** El desarrollo se dividió en tres notebooks:  **1. 01\_eda.ipynb:** Análisis exploratorio de los datos. **2. 02\_modelado.ipynb:** Entrenamiento de modelos K-means y DBSCAN. **3. 03\_visualizacion\_pca.ipynb:** Reducción de dimensionalidad con PCA y visualización de resultados. **Resultados Obtenidos** A continuación, se presentan las visualizaciones de los agrupamientos obtenidos:  📌 **Clustering con K-means (PCA): pca\_kmeans\_clusters.png**    **📌 Clustering con DBSCAN (PCA): pca\_dbscan\_clusters.png**    **📌 Tabla de perfiles por cluster (K-means): tabla\_perfiles\_kmeans.csv** Los resultados muestran tres perfiles bien diferenciados de usuarios en función del total de productos, precio promedio y número de países desde donde compraron.   **Conclusiones**  * El modelo K-means permitió identificar claramente tres perfiles distintos de usuarios. * DBSCAN resultó útil para detectar clientes con comportamiento atípico (outliers). * La visualización con PCA ayudó a comprender la separación entre grupos. * Estas técnicas pueden aplicarse en contextos reales para estrategias de marketing más precisas.  **📚 Estructura del Proyecto** actividad1\_semana3/ ├── data/ → Dataset original y normalizado ├── notebooks/ → Notebooks del desarrollo (EDA, modelado, visualización) ├── results/ → Gráficos y tabla generados ├── docs/ → Informe y presentación final └── README.md → Descripción general para GitHub |