

Proyecto de Minería de Datos – Grupo 4

“Clasificación de clientes según patrones de compra”

Carrera: Ingeniería en Sistemas de Información

Asignatura: Minería de Datos

IV Año – II Semestre

Grupo: 4

Fecha de entrega: **15 de octubre de 2025**

Docente: Verónica Cecilia Morrás Torrez

1. Objetivo del proyecto

Segmentar clientes en clústeres según frecuencia, monto y recencia de compra para proponer estrategias comerciales.

2. Tipo de modelo a utilizar

Clustering con K-Means (y comparación con DBSCAN u otro).

3. Dataset / Recursos

Customer Segmentation Dataset (Kaggle) o datos transaccionales simulados.

4. Fases de desarrollo

Fase	Actividades	Producto
Definición del problema	Formular pregunta, alcance, variables de interés y riesgos.	Ficha de proyecto
Preparación de datos	Recolección/descarga, limpieza, EDA, splits.	Notebook con EDA
Entrenamiento	Aplicar modelo(es), tuning, validación.	Modelo entrenado
Evaluación	Métricas, comparación de modelos, análisis de error.	Informe de resultados
Presentación	Demo funcional + slides de 10 min.	Defensa del proyecto

5. Librerías a instalar

Comando sugerido:

```
pip install matplotlib numpy pandas scikit-learn seaborn
```

Detalle de librerías: matplotlib, numpy, pandas, scikit-learn, seaborn

6. Requisitos técnicos

- Entorno local o Colab.
- Normalización/estandarización; método del codo/silhouette.
- Interpretación de centroides y perfiles de clúster.

Extras opcionales:

- Opcional: app Streamlit para explorar segmentos y filtros.

7. Indicaciones de entrega

- Notebook Jupyter (código comentado reproducible).
- Informe técnico en PDF (Introducción, Dataset, Metodología, Resultados, Conclusiones, Bibliografía APA 7).
- Presentación de máximo 10 minutos (gráficos claros).
- Archivo del modelo entrenado (.pkl, .h5 o similar) y, si aplica, script de inferencia.

8. Rúbrica específica

Criterio	Descripción	Ponderación
Preparación y EDA	Variables RFM, tratamiento de outliers y normalización.	25%
Modelado	Selección K, comparación con otro algoritmo y justificación.	25%
Interpretación	Perfiles de clúster, recomendaciones accionables.	25%
Informe y visuales	Gráficos 2D/3D, claridad narrativa.	15%
Presentación	Comunicación de insights clave.	10%

9. Recomendaciones del docente

- Comenten el código explicando decisiones y supuestos.
- Justifiquen elección del modelo y métricas; incluyan ablation (baseline vs modelo final).
- Eviten plagio y respeten licencias de datos.
- Respalden en Drive o GitHub, controlando versiones.
- Cuiden rutas relativas y orden del notebook (debe correr de arriba a abajo sin errores).
- Incluyan visualizaciones para interpretar resultados (matriz de confusión, curvas ROC/PR, etc.).

Fecha de entrega: 15 de octubre de 2025