

# Rúbrica de Evaluación - Práctica de Minería de Datos

## Proceso KDD con Clustering Jerárquico y K-Means

Calificación Total: 10 puntos

Modalidad: Grupal (2 personas)

Entrega: Un notebook ejecutado por grupo

### 1. SELECCIÓN Y COMPRENSIÓN DE DATOS (1.5 puntos)

Criterio	Excelente (1.5)	Bueno (1.0)	Regular (0.5)	Insuficiente (0.0)
Selección del dataset	Dataset apropiado, relevante y bien justificado. Describe claramente la fuente y características	Dataset apropiado con justificación básica	Dataset poco apropiado o justificación insuficiente	No justifica la selección o dataset inadecuado
Descripción de variables	Descripción completa de todas las variables, tipos de datos y su significado en el contexto	Describe la mayoría de variables adecuadamente	Descripción incompleta o superficial	No describe las variables
Análisis exploratorio inicial	Incluye estadísticas descriptivas, dimensiones y primeras observaciones del dataset	Incluye información básica del dataset	Información mínima e incompleta	No realiza análisis exploratorio

### 2. PREPROCESAMIENTO DE DATOS (2.0 puntos)

Criterio	Excelente (2.0)	Bueno (1.5)	Regular (1.0)	Insuficiente (0.0)
Limpieza de datos	Identifica y trata correctamente valores nulos, duplicados y outliers con justificación	Trata valores nulos y duplicados adecuadamente	Limpieza básica sin justificación clara	No realiza limpieza o es incorrecta
Transformación de datos	Normalización/estandarización apropiada. Codificación correcta de variables categóricas	Aplica transformaciones necesarias correctamente	Transformaciones incompletas o parcialmente correctas	No transforma los datos adecuadamente

Criterio	Excelente (2.0)	Bueno (1.5)	Regular (1.0)	Insuficiente (0.0)
Selección de características	Selecciona y justifica las variables relevantes para clustering. Reduce dimensionalidad si es necesario	Selecciona variables apropiadas con justificación básica	Selección de variables sin justificación clara	No realiza selección de características

### 3. IMPLEMENTACIÓN DE ALGORITMOS (3.5 puntos)

#### 3.1 Clustering Jerárquico Aglomerativo (1.75 puntos)

Criterio	Excelente (1.75)	Bueno (1.25)	Regular (0.75)	Insuficiente (0.0)
Implementación correcta	Implementa correctamente el algoritmo con parámetros apropiados (linkage, métrica de distancia)	Implementa correctamente con parámetros por defecto	Implementación con errores menores	No implementa o tiene errores graves
Dendrograma	Genera y analiza dendrograma para determinar número óptimo de clusters	Genera dendrograma sin análisis profundo	Dendrograma básico o poco claro	No genera dendrograma
Justificación de parámetros	Justifica elección de linkage y métrica de distancia	Menciona parámetros sin justificación detallada	No justifica parámetros	No especifica parámetros

#### 3.2 K-Means (1.75 puntos)

Criterio	Excelente (1.75)	Bueno (1.25)	Regular (0.75)	Insuficiente (0.0)
Implementación correcta	Implementa correctamente K-Means con inicialización apropiada	Implementa correctamente con parámetros por defecto	Implementación con errores menores	No implementa o tiene errores graves
Determinación de K óptimo	Utiliza métodos apropiados (Elbow, Silhouette) para determinar K óptimo con análisis detallado	Utiliza al menos un método para determinar K	Determina K sin justificación adecuada	No determina K óptimo
Comparación de inicializaciones	Compara diferentes inicializaciones o ejecuta múltiples veces	Ejecuta una sola vez adecuadamente	Ejecución básica sin análisis	Implementación deficiente

4. EVALUACIÓN Y ANÁLISIS DE RESULTADOS (2.0 puntos)

Criterio	Excelente (2.0)	Bueno (1.5)	Regular (1.0)	Insuficiente (0.0)
Métricas de evaluación	Calcula y analiza múltiples métricas (Silhouette, Davies-Bouldin, Calinski-Harabasz) para ambos algoritmos	Calcula al menos 2 métricas apropiadamente	Calcula una métrica básica	No evalúa los resultados
Visualización de clusters	Visualizaciones claras, informativas y bien etiquetadas (scatter plots, PCA si es necesario) para ambos algoritmos	Visualizaciones adecuadas para ambos algoritmos	Visualizaciones básicas o incompletas	Visualizaciones ausentes o incorrectas
Interpretación de clusters	Analiza e interpreta el significado de cada cluster, identifica características distintivas	Describe los clusters encontrados	Interpretación superficial	No interpreta los resultados
Comparación de algoritmos	Compara exhaustivamente ambos algoritmos (ventajas, desventajas, resultados obtenidos)	Compara ambos algoritmos básicamente	Comparación mínima	No compara los algoritmos

5. PRESENTACIÓN Y DOCUMENTACIÓN (1.0 punto)

Criterio	Excelente (1.0)	Bueno (0.75)	Regular (0.5)	Insuficiente (0.0)
Organización del notebook	Notebook bien estructurado con secciones claras correspondientes al proceso KDD	Notebook organizado adecuadamente	Organización básica	Notebook desorganizado
Documentación del código	Código bien comentado, con explicaciones claras en markdown entre secciones	Código comentado adecuadamente	Comentarios mínimos	Sin comentarios o explicaciones
Ejecución completa	Notebook ejecutado completamente sin errores, resultados visibles	Notebook ejecutado con errores menores	Notebook parcialmente ejecutado	Notebook no ejecutado o con errores graves
Claridad y redacción	Textos explicativos claros, sin errores ortográficos, lenguaje técnico apropiado	Redacción clara con errores menores	Redacción confusa o con varios errores	Redacción deficiente
Identificación del grupo	Nombres completos de ambos integrantes claramente identificados al inicio	Identificación presente	Identificación incompleta	Sin identificación

## Criterios Adicionales

### Penalizaciones:

- **-1.0 punto:** Entrega fuera del plazo establecido (por día de retraso)
- **-0.5 puntos:** Notebook no ejecutado completamente
- **-0.5 puntos:** Más de un integrante sube el trabajo al aula virtual
- **-2.0 puntos:** Plagio detectado o código copiado sin citar fuentes

### Bonificaciones (máximo +0.5 puntos extra):

- **+0.3 puntos:** Implementa validación cruzada o análisis de estabilidad de clusters
  - **+0.2 puntos:** Utiliza técnicas avanzadas (t-SNE, UMAP para visualización, DBSCAN para comparación adicional)
  - **+0.2 puntos:** Incluye análisis de sensibilidad de parámetros
- 

## Estructura Sugerida del Notebook

1. **Introducción:** Presentación del problema y objetivos
2. **Selección de Datos:** Carga y descripción del dataset
3. **Exploración de Datos:** Análisis exploratorio inicial
4. **Preprocesamiento:** Limpieza, transformación y preparación
5. **Modelado - Jerárquico:** Implementación y análisis
6. **Modelado - K-Means:** Implementación y análisis
7. **Evaluación:** Métricas y comparación
8. **Conclusiones:** Hallazgos principales y recomendaciones