

Universidad Peruana de Ciencias Aplicadas (UPC)

Curso: Data Mining & Data Analysis

Docente: PhD. Carlos Fernando Montoya Cubas

Semestre Académico: 2025-2

TRABAJO FINAL – Minería de Patrones Emergentes en Datos Transaccionales

Modalidad: Grupos de 2 a 3 estudiantes

Exposición Final: Última sesión (10 minutos por grupo)

Valor total: 20 puntos (+1 punto opcional)

INTRODUCCIÓN

El presente trabajo final integra los principales conceptos del curso Data Mining & Data Analysis, aplicando técnicas de minería de patrones frecuentes, análisis comparativo de particiones y detección de patrones emergentes en contextos reales de negocio. El trabajo culmina con una exposición grupal donde los estudiantes presentan e interpretan los resultados obtenidos.

OBJETIVO GENERAL

Evaluar la capacidad del estudiante para aplicar metodologías de minería de datos transaccionales, identificar patrones relevantes, analizar cambios entre subconjuntos mediante Growth Rate, Emerging Patterns (EPs) y Jumping Emerging Patterns (JEPs), y comunicar hallazgos en un escenario de negocio.

ENUNCIADO DEL TRABAJO FINAL

1. Selección y descripción del dataset

- Elegir un dataset transaccional con al menos 1000 transacciones.
- Realizar limpieza y estandarización.
- Describir número de transacciones, ítems y ejemplos de canastas.

2. División del dataset

- Crear al menos dos particiones (temporal, geográfica, estacional o por segmento).
- Justificar la división y explicar expectativas previas.

3. Minería de patrones por partición

- Extraer itemsets frecuentes.
- Generar reglas de asociación (support, confidence, lift).
- Obtener itemsets closed y maximales.

4. Evaluación de patrones emergentes

- Calcular Growth Rate para cada partición.
- Identificar Emerging Patterns (EPs).
- Identificar Jumping Emerging Patterns (JEPs).
- Interpretar cambios entre divisiones.

5. Interpretación de resultados

- Comparar patrones estables, crecientes y decrecientes.
- Describir implicancias de negocio.
- Explicar hallazgos relevantes.

6. Opcional: SkyPatterns

- Incorporar una métrica personalizada (rentabilidad, utilidad, riesgo, etc.).
- Identificar patrones dominantes bajo la métrica.

7. Entregables

- A. Notebook (Jupyter/Colab) con desarrollo técnico completo.
- B. Informe escrito (máx. 6 páginas) centrado en interpretación.
- C. Exposición final (10 minutos) durante la última sesión.

RÚBRICA DEL TRABAJO FINAL (Total: 20 puntos)

1. Preparación y división del dataset – 4 puntos

Selección adecuada, limpieza apropiada, descripción clara y justificación sólida de las particiones.

2. Minería de patrones – 4 puntos

Generación correcta de itemsets frecuentes, reglas de asociación, itemsets closed y maximales.

3. Growth Rate, Emerging Patterns y Jumping Emerging Patterns – 4 puntos

Cálculo adecuado de Growth Rate e identificación precisa de EPs y JEPs, con análisis comparativo.

4. Interpretación y conclusiones de negocio – 4 puntos

Explicaciones claras, análisis profundo, coherencia entre hallazgos y conclusiones de negocio.

5. Exposición final – 4 puntos

Claridad, organización, dominio del tema, uso de visualizaciones y capacidad de síntesis.

BONUS OPCIONAL

Aplicación de métricas personalizadas mediante SkyPatterns: +1 punto adicional.

FIN DEL DOCUMENTO