

Minería de Datos - FP-Growth (INE 2024)

Alumno: David Barral

Fecha: 18/10/2025

Objetivo

Aplicar el algoritmo FP-Growth para descubrir reglas de asociación en la base de violencia intrafamiliar 2024. Se presentan tres patrones con su interpretación y métricas.

Alcance y Nota Importante

El análisis se realizó a nivel nacional, sin aplicar los filtros de la Sesión 3. Se discretizaron variables numéricas con `arules::discretize(method='frequency', categories=5)` y se estimaron las reglas mediante FP-Growth usando `mlxtend` vía `reticulate`. Métricas: soporte, confianza y lift (umbrales: `min_support=0.20`, `min_confidence=0.50`).

1. Patrón - Víctima (Nacional)

Regla: `VIC_DEDICA = [1,9] ⇒ VIC_TRABAJA = [2,9]`. Métricas: Soporte ~ 0.615, Confianza = 1.00, Lift ~ 1.627. Interpretación: ciertas categorías de dedicación de la víctima se asocian de forma determinista con valores de 'trabaja' en [2,9]; el `lift>1` indica co-ocurrencia por encima de lo esperado.

2. Patrón - Víctima (Nacional)

Regla: `VIC_TRABAJA = [2,9] ⇒ VIC_DEDICA = [1,9]`. Métricas: Soporte ~ 0.615, Confianza = 1.00, Lift ~ 1.627. Interpretación: relación recíproca del patrón anterior; evidencia una alineación alta entre condición laboral y dedicación tras la discretización (no implica causalidad).

3. Patrón - Agresor (Nacional)

Regla: `AGR_TRABAJA = [2,9] ⇒ AGR_DEDICA = [1,9]`. Métricas: Soporte ~ 0.317, Confianza = 1.00, Lift ~ 3.151. Interpretación: ambas condiciones coexisten con probabilidad muy superior a la esperada; sugiere perfiles ocupacionales con baja inserción formal.

Limitaciones

Los intervalos como [1,9] provienen de la discretización (quintiles por frecuencia) y/o codificación del instrumento; representan clases, no magnitudes exactas. La confianza 1.00 no implica causalidad. Cambiar support (p. ej., 0.10) puede revelar más reglas; validar con contexto.

Reproducibilidad

El script `run_fpgrowth.R` lee el Excel local, discretiza, transforma a one-hot y ejecuta FP-Growth con `mlxtend`. Exporta CSVs en `./out/` y ordena reglas por lift, luego confianza y soporte.