

Predicción de Zonas de Violencia Intrafamiliar con Random Forest

Análisis predictivo de la ubicación geográfica de hechos de violencia intrafamiliar en 2024 mediante técnicas de aprendizaje automático.

Preparación de los Datos

VARIABLES SELECCIONADAS

De la víctima: sexo, edad, escolaridad, situación laboral y dedicación

Del agresor: sexo, edad, escolaridad y situación laboral

Variable objetivo: zona del hecho (urbana, rural u otras áreas)

PROCESO DE LIMPIEZA

- Eliminación de valores faltantes
- Conversión de variables categóricas
- Aleatorización de registros
- División: 80% entrenamiento, 20% prueba

Modelo Random Forest



200 Árboles

Combinación de múltiples árboles de decisión para mejorar la precisión



mtry = 9

Variables evaluadas en cada división del árbol



Error OOB: 47%

Rendimiento razonable para clasificación de tres clases con datos reales

El Random Forest supera a un árbol de decisión tradicional al combinar resultados de múltiples árboles, reduciendo el sobreajuste y mejorando la generalización.

Resultados del Modelo

53.56%

Precisión Global

Superior al azar (33%) en problema de
3 clases

1206

Aciertos Zona Urbana

Clase mejor predicha del modelo

47%

Tasa de Error

Estimación OOB del error promedio

Matriz de Confusión

Real / Predicción	Urbana (1)	Rural (2)	Otras (9)
Urbana (1)	1206	862	51
Rural (2)	869	1184	39
Otras (9)	156	113	21

La zona urbana es la mejor predicha. Existe confusión entre zonas urbanas y rurales debido a características similares. La clase "otras áreas" es la más difícil por su menor representación en los datos.

Evolución del Error del Bosque

01

Estabilización en árbol 150

El error del modelo se estabiliza alrededor del árbol número 150

02

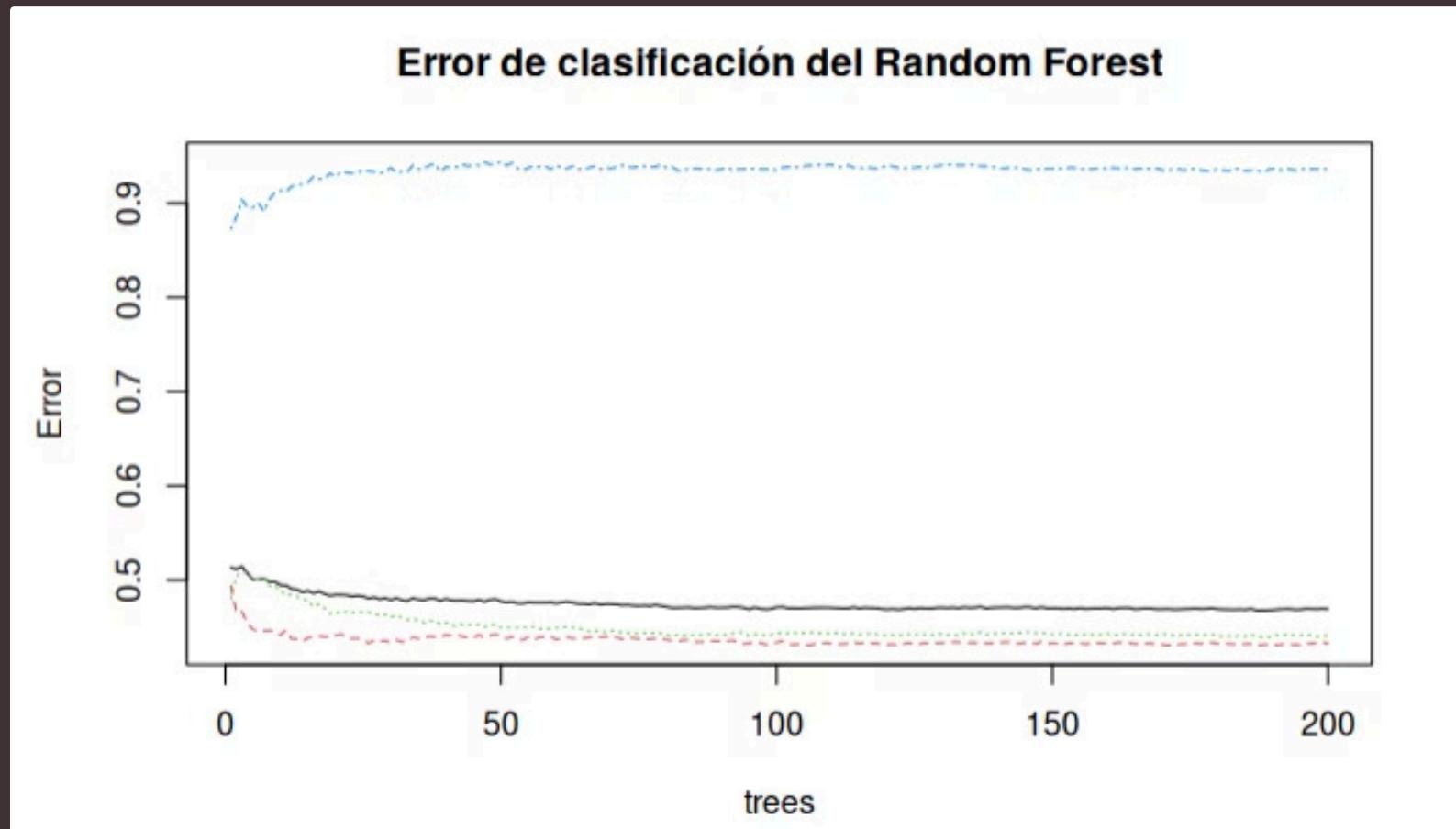
Clase 1 con menor error

La zona urbana presenta la línea de error más baja

03

Clase 9 más difícil

Otras áreas muestran el error más alto por menor cantidad de datos



Predicción de Caso Simulado



Perfil Analizado

Víctima

Mujer, 25 años, escolaridad media, estudiante y trabajadora

Agresor

Hombre, 30 años, escolaridad universitaria

Predicción: Zona Urbana

Mayor probabilidad de ocurrencia en área urbana

La predicción coincide con la tendencia observada: personas con mayor escolaridad y actividad económica suelen ubicarse en áreas urbanas.

Conclusiones

Rendimiento Mejorado

Random Forest supera a un árbol de decisión individual con precisión del 53.56%

Variables Influyentes

Escolaridad, edad, dedicación y ocupación son los factores más determinantes

Patrones Identificados

Zona urbana más predecible; "otras áreas" difícil por menor tamaño muestral

Aplicación Práctica

Permite estudiar distribución espacial según características sociodemográficas

Reflexión Final



Identificar Patrones

Reconocer distribución geográfica de la violencia



Perfiles de Riesgo

Detectar características asociadas a mayor vulnerabilidad



Políticas Públicas

Apoyar decisiones basadas en evidencia

La minería de datos permite analizar fenómenos sociales complejos como la violencia intrafamiliar, mejorando la focalización de programas de prevención, atención y denuncia.