

PROYECTO DE APRENDIZAJE AUTOMÁTICO

Tema: Detección temprana de situaciones de trata de personas (foco en Tierra del Fuego)

Modelado supervisado y priorización con umbral por PR (Argentina 2020–2024)

Resumen ejecutivo

Se desarrolló un sistema para priorizar posibles casos de trata a partir de registros 2020–2024 (**7.848 filas, 26 columnas**). Se construyó robustamente la etiqueta `es_trata` y se aplicó un flujo con **split temporal**, **búsqueda de umbral** por curva **Precision–Recall**, **calibración** de probabilidades, **backtesting** rolling-origin e **interpretabilidad** (Permutation Importance y SHAP).

Propuesta operativa.

Se recomienda: **Logistic Regression con interacciones + umbral 0.345** (elegido por PR con recall ≥ 0.80).

En test: Precision 0.563 · Recall 0.972 · F1 0.713 · ROC-AUC 0.623.

Alternativa si se prioriza F1/ROC: **HistGradientBoosting calibrado @0.396 (Precision 0.562 · Recall 0.958 · F1 0.708 · ROC-AUC 0.659 · AP 0.685).**

El **backtesting mensual** muestra promedios: **P 0.647 · R 0.686 · F1 0.651.**

Se adjunta pipeline y plantilla de inferencia.

Problema, objetivo y alcance

- **Objetivo:** priorizar intervenciones con mayor probabilidad de corresponder a trata (`es_trata=1`).
- **Uso esperado:** ranking/priorización para equipos de respuesta temprana.
- **Alcance:** todo el país (TDF como foco exploratorio, no exclusivo).

- **Supuestos clave:** consistencia temporal de criterios de carga; sesgos operativos mitigados vía calibración y monitoreo.

Datos y preparación

- **Forma final:** 7.848 registros · 26 variables (ver results/ para perfiles).
- **Nulos (top):** p.ej. consultante_como_conocio_la_linea 6.806; fechas derivadas sin nulos operativos.
- **Balance:** es_trata=1 **54% (4.241)** / 0 **46% (3.607)**.
- **Limpieza:** estandarización de categorías/fechas; **eliminación de 5 filas** sin etiqueta válida.
- **Ingeniería:** derivaciones de año/mes/día/semana/trimestre, **mes seno/cos**, banderas (es_fin_semana, es_anonima), IDs geográficos y orígenes de institución.

Cambios aplicados a la base original (trazabilidad)

Este bloque documenta explícitamente las modificaciones que se hicieron sobre el dataset bruto previo al modelado.

Modificaciones manuales declaradas por el equipo

- Eliminación de columna: hora_ingreso (vacía).
- Construcción de etiqueta binaria es_trata (reglas auditadas; ver Sección 3).

Compleción de columnas casi vacías (detallado por campo)

- consultante_provincia: (1) mapear por provincia_indec_id; (2) normalizar; (3) corregir variantes (p.ej. CABA ↔ Ciudad Autónoma de Buenos Aires); (4) si no hay fuente confiable → sin_descripcion.
Ej.: “C.A.B.A.” → “ciudad autonoma de buenos aires”.
- consultante_localidad: estandarizar y homogeneizar; si no se recupera → sin_descripcion.
Ej.: “Río Gránde” → “rio grande”.
- consultante_nacionalidad: consolidar sinónimos/errores; minúsculas y sin tildes; sin dato → sin_descripcion.
Ej.: “BOLÍVIA” → “bolivia”.

Controles y trazabilidad

Tablas de frecuencias antes/después y **reporte de nulos** en results/. En el pipeline, imputaciones ajustadas **sólo con TRAIN**.

Otras transformaciones

Normalización de strings; homogeneización geográfica; derivaciones temporales y componentes cíclicos; banderas; uso de IDs.

Impacto y control de fuga

Decisiones de eliminación/imputación **sin fuga** (antes del split o fit-only-train). Snapshot del bruto + notebook reproducible + tablas de nulos.

QA y construcción de es_trata

- Chequeos de consistencia y distribución temporal/espacial.
- Respaldo visual en figs: barras por año/mes/trimestre y heatmaps provincia×trimestre.

EDA temporal (2020–2024)

- **Intervenciones por año:** 2020→2024 con crecimiento, **pico 2024 (~769)**.
- **Tasa anual** de trata ~**0.48–0.56** (estable/ligero aumento).
- Serie mensual final exportada: outputs/serie_mensual_final.csv.
 - Figuras: figs/serie_mensual_final_total.png, figs/serie_mensual_final_positivos.png, figs/serie_mensual_final_tasa.png.
- Backups: tablas mensuales y trimestrales en results/.

EDA espacial (Top-K)

- **Provincias con más intervenciones** (Top): Buenos Aires, CABA, Santa Fe, Córdoba, Tucumán, etc.
- **Top por tasa:** La Rioja, Tierra del Fuego, San Luis, San Juan... (cautela por N).
- **Localidades:** concentración en **Buenos Aires** y **CABA**; ranking y tasas (mín. 5 casos).
- **Nacionalidades:** mayoría **Argentina**; ranking de tasas con N mínimo.

Pruebas χ^2 (variable × es_trata)

consultante_nacionalidad $\chi^2=148.79$ ($p<1e-22$); consultante_provincia $\chi^2=126.01$ ($p<1e-15$);
consultante_localidad $\chi^2=499.58$ ($p<1e-8$); es_anonima $\chi^2=46.51$ ($p<1e-11$);
mes $\chi^2=20.74$ ($p\approx 0.036$); temporada $\chi^2=7.34$ ($p\approx 0.119$).

Tablas con %fila/%col en anexos.

Metodología de validación y modelado

- **Split temporal (sin fugas):** train/valid/test por fechas; transformaciones **fit sólo en train**.
- **Optimización del umbral:** por **curva Precision–Recall**, con restricción **recall ≥ 0.80** .
- **Calibración:** Isotónica/Platt según modelo; evaluación por **Brier score** y curva de calibración.
- **Backtesting temporal (rolling origin mensual):** 2020-07→2024-12; reporte por mes y promedio final.
- **Pipelines** persistidos y reproducibles; semillas fijas.

Modelos evaluados (resumen)

- **Logistic Regression (base) @0.328 (PR):** P 0.559 • R 0.951 • F1 0.704 • AP 0.667 • ROC-AUC 0.628.
- **Logistic Regression + interacciones** (temporada×anonimato, provincia×anonimato, nacionalidad×temporada) @0.345:
P 0.563 • R 0.972 • F1 0.713 • ROC-AUC 0.623.

- **HistGradientBoosting (calibrado) @0.396: P 0.562 · R 0.958 · F1 0.708 · ROC-AUC 0.659 · AP 0.685.**

Notas: RandomForest simple descartado (recall pos=0 @0.50 en test pequeño). Comparativas de **CV/Holdout y tuning** en results/ (ROC/PR/CM).

Calibración de probabilidades

- Mejora de **Brier** (p.ej., HGB: 0.243→0.234) y alineamiento visual tras calibración.
- **Umbral operativo recomendado: 0.345** (LogReg+interacciones) · **alternativo 0.396** (HGB calibrado).
- **Persistencia:** results/mejor_pipeline_calibrado.pkl + results/best_threshold_calibrado.json.

Backtesting temporal (rolling origin)

- **Promedios: Precision 0.647 · Recall 0.686 · F1 0.651.**
- Tabla completa por período en notebook; últimos meses 2024-11/12 con F1 alto (≈0.72–0.78).

Interpretabilidad

- **Permutation Importance (HGB):** variables temporales (**trimestre**, **mes_cos**), **origen_institucion** y señales operativas.
- **SHAP:** destaca **edad_aparente**, componentes temporales y ubicación.
Figuras y CSV en results/ y figs/.

Foco Tierra del Fuego (TDF)

- **n=13 registros** (tablas en results/ y gráficos TDF en figs/).
- Con **LogReg+interacciones y mismo umbral (0.345): n=30 (corrida específica) → Precision 0.733 · Recall 1.00 · F1 0.846** — interpretar con cautela por bajo N.
- **Recomendación:** usar **modelo general** con **recalibración provincial** y monitoreo.

Inferencia

- Plantilla con columnas esperadas: results/inferencia_template_columns.csv.
- Demo desde X_test: results/inferencia_DEMO_desde_Xtest.csv (idx, proba, pred_label).
- **Umbral aplicado según versión: 0.345 (LogReg+interacciones) o 0.396 (HGB calibrado).**

Recomendación final

- **Modelo seleccionado para operación: Logistic Regression + interacciones @0.345 (PR).**
- **Máximo Recall (0.972)** con **buena Precision (0.563)** y **F1 (0.713)**. Simple, transparente y barato de mantener.
- **Plan B** (si se prioriza F1/ROC y calibración): **HGB calibrado @0.396.**

Riesgos, sesgos y mitigaciones

- Sesgo de reporte: mitigar con calibración periódica.

- Drift temporal/espacial: monitorear tasa de positivos, Brier, PR/F1 por mes y provincia; gatillar recalibración si $|\Delta \text{Brier}| > 0.03$ o caída de recall > 5 pp.
- Clases minoritarias por localidad/nacionalidad: umbrales mínimos de N para tasas; evitar conclusiones por microgrupos.

Próximos pasos

1. Monitoreo mensual con tablero (series tot/pos/tasa + métricas).
2. Recalibración trimestral y por provincia si cambia la casuística.
3. Evaluar lags de serie y features operativas adicionales.
4. Integrar explicaciones locales (SHAP force) en la vista operativa.

Entregables

- **Figuras:** ver carpeta figs/ (PR/ROC/CM, heatmaps, Top-K, TDF, series mensuales).
- **Resultados:** métricas, umbrales y tablas en results/.
- **Series:** outputs/serie_mensual_final.csv.
- **Pipelines:** results/mejor_pipeline.pkl, results/mejor_pipeline_calibrado.pkl.
- **Inferencia demo/plantilla:** results/inferencia_DEMO_desde_Xtest.csv, results/inferencia_template_columns.csv.
- **Contacto / notas:** el repo/notebook incluye **semillas y paths relativos** para reproducción total. Anexos con tablas completas disponibles si se requiere.

Contacto / notas: el repo/notebook incluye semillas y paths relativos para reproducción total. Sección de anexos con tablas completas disponible si se requiere.

Anexo A – Trazabilidad de limpieza (antes → después)

Mini-tablas ilustrativas con ejemplos de normalización y referencia a los CSV de auditoría en results/.

A.1 Provincias (mapeo desde provincia_indec_id + normalización)

Ejemplos de corrección

Antes (raw)	Después (normalizado)	Regla aplicada
C.A.B.A.	ciudad autonoma de buenos aires	alias → nombre INDEC
Bs As	buenos aires	alias → nombre INDEC
Neuquén	neuquen	minúsculas + sin tildes

Archivos de respaldo: results/auditoria_provincia_antes.csv, results/auditoria_provincia_despues.csv (frecuencias).

A.2 Localidades (estandarización y fallback)

Ejemplos de corrección

Antes (raw)	Provincia	Después (normalizado)	Regla aplicada
Río Gránde	tierra del fuego	rio grande	quitar tildes + trim
CABA	buenos aires	caba	alias estándar
(vacío)	buenos aires	sin_descripcion	sin fuente confiable

Archivos de respaldo: results/auditoria_localidad_antes.csv, results/auditoria_localidad_despues.csv.

A.3 Nacionalidades (consolidación y categorías)

Ejemplos de corrección

Antes (raw)	Después (normalizado)	Regla aplicada
ARG / Argentina	argentina	consolidar sinónimos
BOLÍVIA	bolivia	quitar tildes
(vacío)	sin_descripcion	sin fuente confiable

Archivos de respaldo: results/auditoria_nacionalidad_antes.csv, results/auditoria_nacionalidad_despues.csv.

A.4 Resumen de nulos (antes → después)

- Bajaron nulos en campos clave:
 - *consultante_localidad*: 7079 → 6389 (−690; −8,8 pp)
 - *consultante_nacionalidad*: 6464 → 5940 (−524; −6,7 pp)
 - *consultante_provincia*: 3737 → 3154 (−583; −7,5 pp)
- *hora_ingreso*: columna eliminada por vacía (ruido).
- *via_ingreso*: ahora visible con 617 nulos (7,86%).
- Resto estable; sin deterioro en variables operativas.

Interpretación breve. Se reducen nulos donde más importa (territorio/identificación), se preserva la trazabilidad y el dataset queda listo para el split temporal, calibración y modelado.

Archivos de respaldo: results/nulos_antes.csv, results/nulos_despues.csv.

Nota: Estas mini–tablas son ilustrativas; los conteos completos y porcentajes están disponibles en los CSV indicados en results/.