



PROYECTO FINAL - APRENDIZAJE AUTOMÁTICO

CLASIFICACIÓN BINARIA DE SISMOS EN TIERRA DEL FUEGO

Presentación realizada por Valeria Villegas



¿POR QUÉ ESTE PROYECTO?

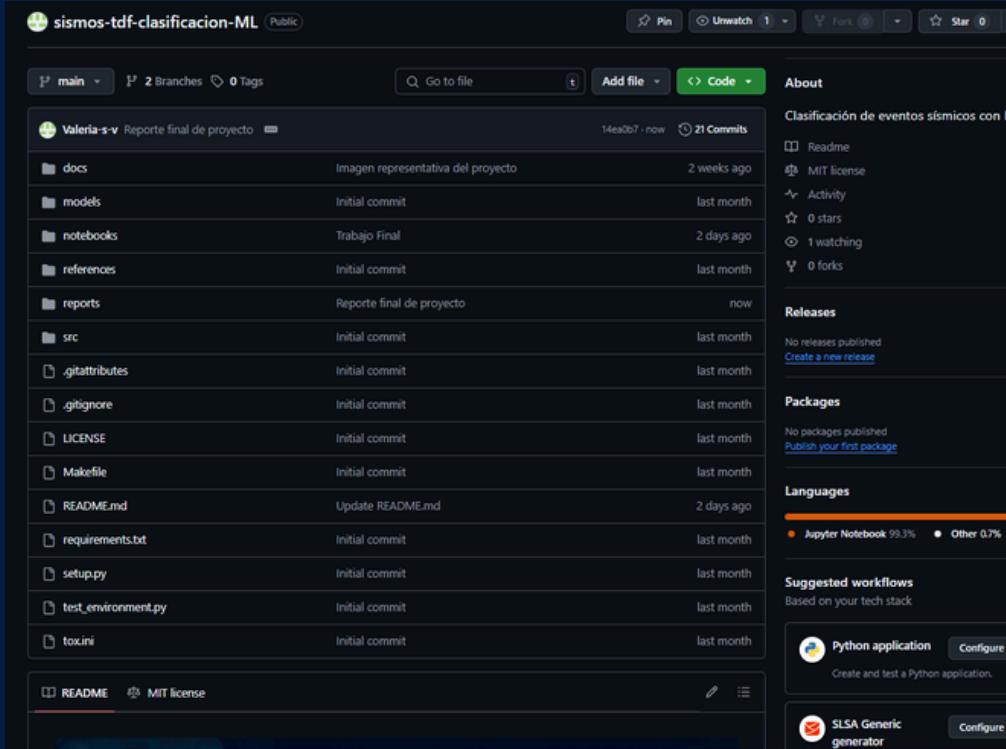
Tierra del Fuego ha sido clasificada por el INPRES como zona de alto riesgo sísmico. Eventos como el sismo del 2 de mayo, percibido en toda la isla y el sur de Chile, refuerzan la urgencia de fortalecer la preparación e información a la población.

01 El objetivo fue construir un modelo de clasificación binaria que permita predecir si un evento sísmico en Tierra del Fuego puede ser percibido por la población.

02 Esta información resulta clave para fortalecer la gestión del riesgo, ya que, aunque los sismos no son frecuentes, pueden generar alarma o causar daño.

03 Además, su aplicación aporta valor a la planificación preventiva, considerando que un evento poco frecuente puede tener un impacto significativo a nivel social y estructural.

ESTRUCTURA Y HERRAMIENTAS



Trabajo en entorno Jupyter y Colab con Python

JUPYTER COLAB

Limpieza y preparación de datos,
EDA, modelado y evaluación.

4 ETAPAS

GITHUB

Repositorio Git estructurado con
CookieCutter

LIBRERÍAS

Uso de librerías: pandas, scikit-learn,
matplotlib

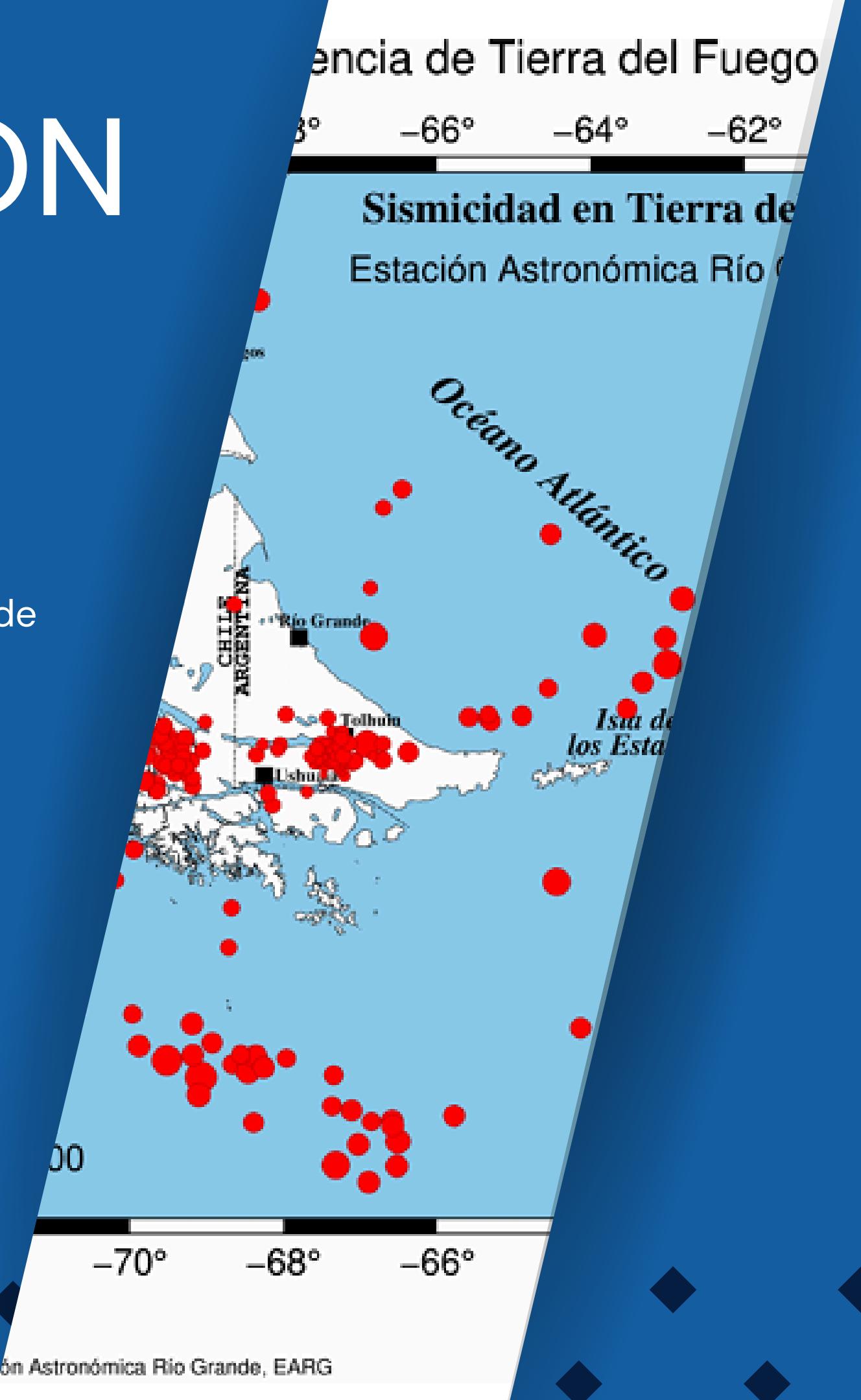
Estructura del proyecto

```
notebooks/          # Notebook principal con workflow completo
src/               # Módulos en Python (opcional)
models/            # Modelos serializados (joblib/pickle)
docs/              # Imágenes, gráficos, documentación visual
reports/           # Exportaciones PDF de informes
references/        # Fuentes & datos originales
requirements.txt   # Dependencias necesarias
README.md          # Este archivo ✨
LICENSE            # Licencia del proyecto (MIT)
```

ORIGEN Y PREPARACIÓN DE LOS DATOS

| # | Año | M | D | H | M | S | Err. | L | Latitud | Err. | Longitud | Err. | Prof. | Err. | NST | RMS | GAP | ML | Agen. |
|-----|------|----|----|----|----|------|------|---|---------|------|----------|-------|-------|-------|-----|-----|-----|-----|-------|
| 205 | 2023 | 7 | 30 | 17 | 2 | 35.2 | 2.53 | R | -53.327 | 46.2 | -62.269 | 42.0 | 0.8 | 23.7 | 5 | 1.2 | 182 | 3.8 | EAR |
| 206 | 2023 | 8 | 4 | 12 | 41 | 53.6 | 0.52 | L | -54.367 | 4.8 | -67.423 | 10.0 | 57.0 | 6.5 | 3 | 0.2 | 222 | 2.2 | EAR |
| 207 | 2023 | 8 | 10 | 6 | 45 | 34.3 | 0.68 | R | -54.666 | 3.5 | -69.306 | 14.6 | 7.7 | 4.1 | 3 | 0.3 | 295 | 2.3 | EAR |
| 208 | 2023 | 8 | 30 | 5 | 51 | 9.9 | 1.40 | L | -54.738 | 6.6 | -69.265 | 13.0 | 6.4 | 15.9 | 5 | 0.5 | 259 | 2.3 | EAR |
| 209 | 2023 | 9 | 7 | 23 | 37 | 31.0 | 1.01 | R | -52.585 | 11.8 | -72.700 | 34.3 | 10.0 | 7.2 | 7 | 0.4 | 239 | 3.1 | EAR |
| 210 | 2023 | 9 | 10 | 7 | 55 | 57.1 | 3.26 | R | -54.728 | 14.6 | -69.537 | 28.0 | 5.8 | 18.4 | 5 | 1.3 | 265 | 2.2 | EAR |
| 211 | 2023 | 9 | 11 | 11 | 3 | 40.5 | 0.44 | R | -54.726 | 3.4 | -69.362 | 5.6 | 9.5 | 4.1 | 5 | 0.1 | 261 | 2.3 | EAR |
| 212 | 2023 | 9 | 14 | 19 | 51 | 58.2 | 0.45 | L | -54.428 | 3.4 | -69.554 | 3.9 | 3.0 | 2.0 | 6 | 0.1 | 202 | 2.4 | EAR |
| 213 | 2023 | 10 | 16 | 9 | 32 | 30.0 | 0.29 | L | -54.658 | 0.8 | -67.223 | 4.9 | 13.5 | 1.5 | 3 | 0.0 | 262 | 2.4 | EAR |
| 214 | 2023 | 10 | 30 | 21 | 1 | 44.1 | 0.51 | R | -50.758 | 5.6 | -74.402 | 9.9 | 10.8 | 2.3 | 6 | 0.2 | 317 | 3.7 | EAR |
| 215 | 2023 | 11 | 26 | 12 | 21 | 46.9 | 1.13 | R | -53.452 | 17.0 | -62.805 | 15.0 | 3.1 | 5.4 | 5 | 0.5 | 323 | 3.5 | EAR |
| 216 | 2023 | 11 | 26 | 14 | 36 | 51.9 | 1.04 | R | -54.151 | 16.9 | -61.172 | 374.2 | 12.9 | 201.7 | 5 | 0.5 | 333 | 3.4 | EAR |
| 217 | 2023 | 11 | 27 | 6 | 40 | 22.9 | 0.57 | R | -53.495 | 10.3 | -68.641 | 4.1 | 3.4 | 4.6 | 3 | 0.2 | 257 | 2.2 | EAR |
| 218 | 2024 | 1 | 13 | 16 | 28 | 58.9 | 1.28 | R | -54.736 | 19.8 | -58.424 | 17.1 | 4.8 | 6.8 | 7 | 0.6 | 263 | 3.7 | EAR |
| 219 | 2024 | 2 | 7 | 1 | 41 | 59.6 | 1.11 | L | -54.616 | 5.4 | -69.053 | 13.3 | 7.2 | 12.9 | 4 | 0.4 | 218 | 2.3 | EAR |
| 220 | 2024 | 2 | 27 | 6 | 23 | 60.0 | 1.05 | L | -56.027 | 14.2 | -71.073 | 9.8 | 20.0 | 8.1 | 7 | 0.5 | 289 | 2.8 | EAR |
| 221 | 2024 | 3 | 11 | 13 | 6 | 53.2 | 1.89 | L | -53.739 | 15.3 | -71.785 | 22.8 | 13.7 | 10.0 | 5 | 0.8 | 257 | 2.8 | EAR |
| 222 | 2024 | 3 | 14 | 13 | 13 | 7.0 | 0.51 | L | -52.970 | 7.2 | -72.875 | 27.1 | 5.0 | 11.6 | 6 | 0.2 | 320 | 3.3 | EAR |
| 223 | 2024 | 3 | 26 | 2 | 16 | 57.2 | 0.34 | L | -57.331 | 5.4 | -68.388 | 15.4 | 11.9 | 2.8 | 3 | 0.1 | 328 | 2.8 | EAR |
| 224 | 2024 | 4 | 9 | 11 | 26 | 59.2 | 0.45 | L | -53.929 | 14.6 | -71.009 | 23.4 | 20.4 | 7.6 | 5 | 0.1 | 232 | 2.8 | EAR |
| 225 | 2024 | 4 | 14 | 3 | 39 | 46.7 | 0.38 | R | -54.562 | 2.0 | -67.144 | 12.1 | 12.4 | 3.1 | 4 | 0.1 | 265 | 2.3 | EAR |
| 226 | 2024 | 4 | 14 | 16 | 44 | 2.3 | 1.23 | R | -54.054 | 8.8 | -70.857 | 13.1 | 5.9 | 4.8 | 5 | 0.5 | 232 | 2.6 | EAR |
| 227 | 2024 | 4 | 18 | 13 | 15 | 33.7 | 0.94 | R | -54.297 | 17.3 | -63.533 | 13.4 | 13.7 | 5.7 | 4 | 0.4 | 335 | 2.8 | EAR |
| 228 | 2024 | 4 | 19 | 10 | 48 | 23.1 | 0.46 | R | -56.628 | 11.0 | -69.189 | 27.1 | 11.5 | 3.6 | 3 | 0.2 | 343 | 3.1 | EAR |
| 229 | 2024 | 4 | 20 | 9 | 24 | 12.7 | 0.84 | R | -56.941 | 8.0 | -68.259 | 34.1 | 22.0 | 8.0 | 2 | 0.2 | 347 | 3.1 | EAR |
| 230 | 2024 | 5 | 12 | 15 | 39 | 12.2 | 0.40 | R | -54.351 | 7.5 | -64.894 | 11.5 | 14.1 | 3.7 | 5 | 0.1 | 309 | 2.8 | EAR |
| 231 | 2024 | 5 | 15 | 14 | 56 | 13.7 | 0.85 | R | -57.630 | 39.0 | -67.318 | 27.7 | 30.5 | 39.1 | 6 | 0.4 | 288 | 3.9 | EAR |
| 232 | 2024 | 5 | 29 | 13 | 23 | 44.4 | 1.37 | R | -50.278 | 18.5 | -71.917 | 53.0 | 0.0 | 12.6 | 4 | 0.6 | 332 | 3.5 | EAR |
| 233 | 2024 | 6 | 16 | 20 | 19 | 28.8 | 0.88 | R | -57.636 | 6.5 | -66.523 | 13.7 | 5.0 | 4.6 | 5 | 0.4 | 285 | 3.2 | EAR |
| 234 | 2024 | 6 | 19 | 18 | 10 | 47.2 | 0.34 | R | -55.582 | 20.2 | -70.191 | 19.2 | 22.2 | 16.6 | 3 | 0.0 | 293 | 2.4 | EAR |
| 235 | 2024 | 6 | 24 | 12 | 24 | 44.1 | 0.30 | L | -54.661 | 4.2 | -67.648 | 5.0 | 10.0F | 0.0 | 3 | 0.1 | 236 | 1.9 | EAR |
| 236 | 2024 | 6 | 29 | 8 | 25 | 18.1 | 0.96 | L | -54.524 | 3.4 | -67.246 | 19.8 | 10.0F | 0.0 | 3 | 0.5 | 234 | 2.2 | EAR |
| 237 | 2024 | 8 | 21 | 2 | 13 | 14.6 | 0.42 | R | -53.687 | 9.2 | -73.840 | 81.0 | 31.5 | 42.9 | 5 | 0.1 | 312 | 3.2 | EAR |
| 238 | 2024 | 8 | 23 | 5 | 53 | 0.5 | 1.53 | R | -53.793 | 67.6 | -59.306 | 35.3 | 32.3 | 74.0 | 5 | 0.7 | 338 | 3.4 | EAR |
| 239 | 2024 | 9 | 15 | 7 | 51 | 53.1 | 0.93 | R | -52.263 | 17.4 | -72.090 | 70.2 | 13.4 | 6.1 | 4 | 0.4 | 209 | 3.2 | EAR |
| 240 | 2024 | 11 | 6 | 13 | 32 | 18.2 | 1.23 | L | -54.603 | 1.1 | -67.563 | 1.8 | 6.0 | 3.6 | 4 | 0.1 | 154 | 2.2 | EAR |

- Fuente: Catálogo Sismológico de Referencia de Tierra del Fuego (2021-2024)
- Tipo de datos: Object, float, entero
- Instancias: 240
- Columnas: 20



PREPARACIÓN Y ENRIQUECIMIENTO DE DATOS

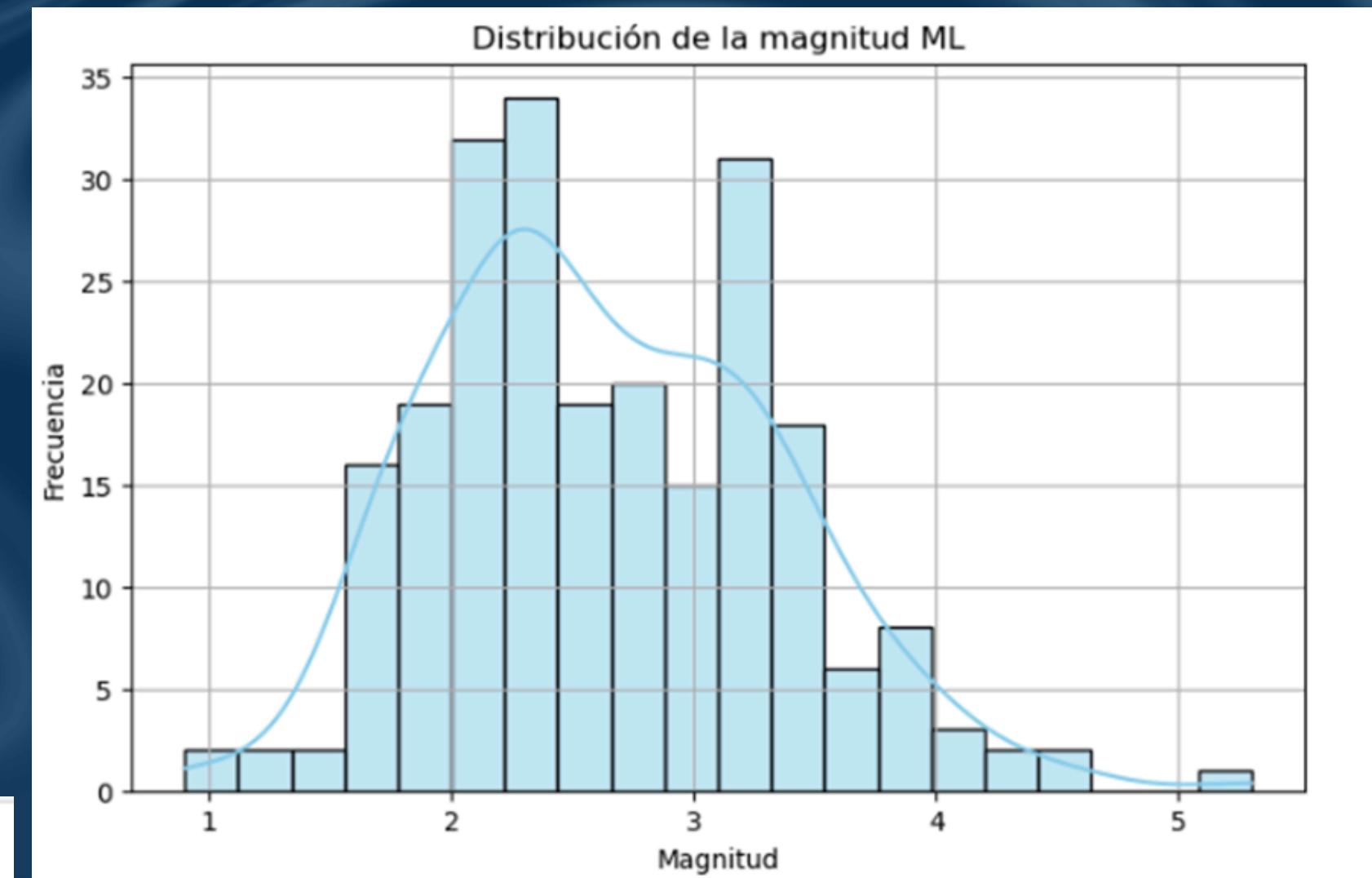
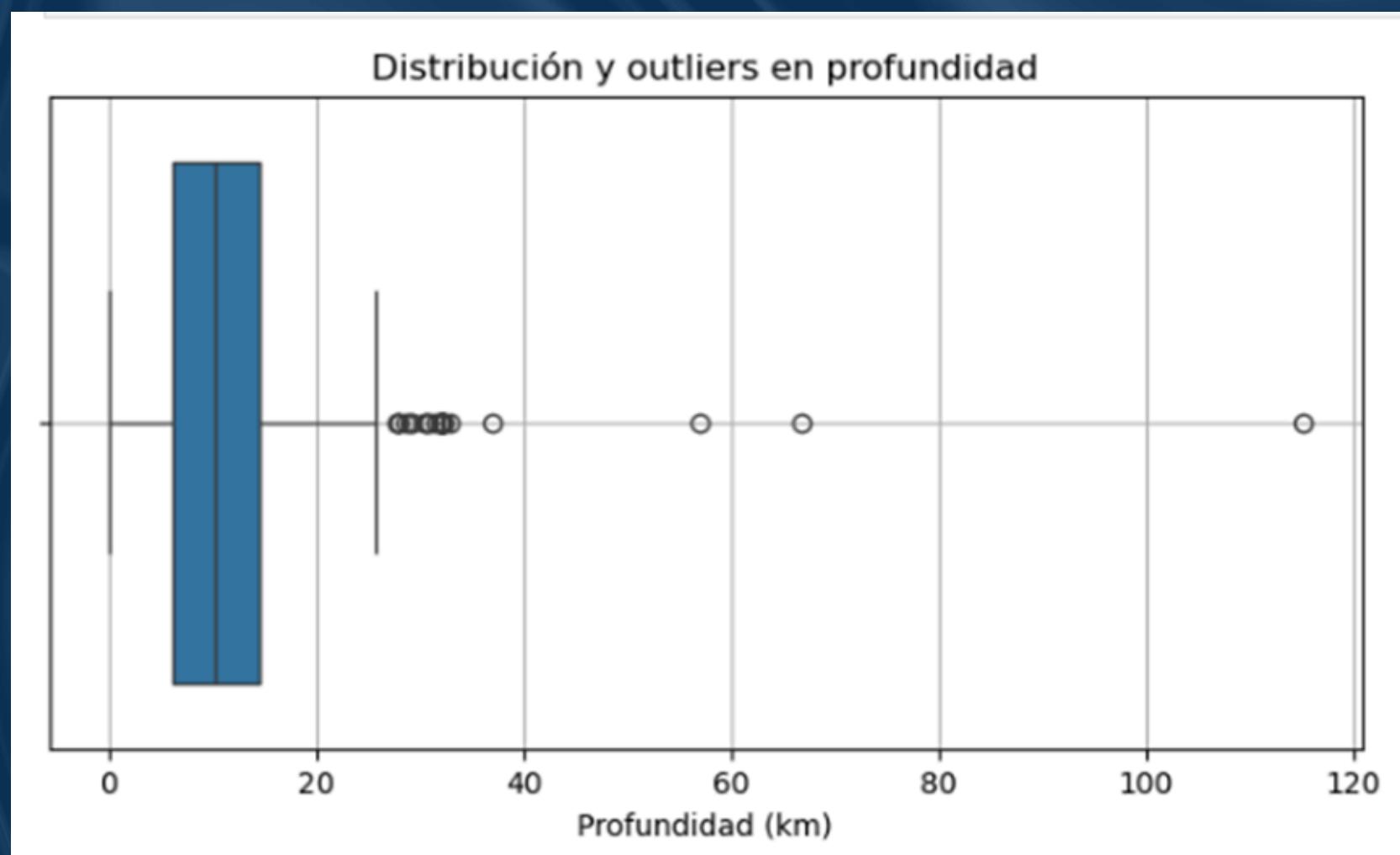
¿CÓMO SE CONSTRUYÓ LA VARIABLE OBJETIVO?

- Se creó una variable binaria llamada “percibido”, con base en dos criterios:
 - Magnitud mayor o igual a 3.5
 - Distancia mínima al centro urbano menor o igual a 250 km
- Esta aproximación busca representar si un sismo podría ser detectado por la población.
- Se midió qué tan lejos estuvo el epicentro de cada una de las tres ciudades principales de Tierra del Fuego: Ushuaia, Río Grande y Tolhuin.

ANÁLISIS EXPLORATORIO

¿QUÉ NOS DICEN LOS DATOS?

- La mayoría de los eventos tienen magnitud entre 2.0 y 3.9 → no suelen percibirse.
- La mayoría son sismos superficiales (<20 km), aunque hay outliers >60 km.
- La variable percibido está desbalanceada → más eventos negativos que positivos.

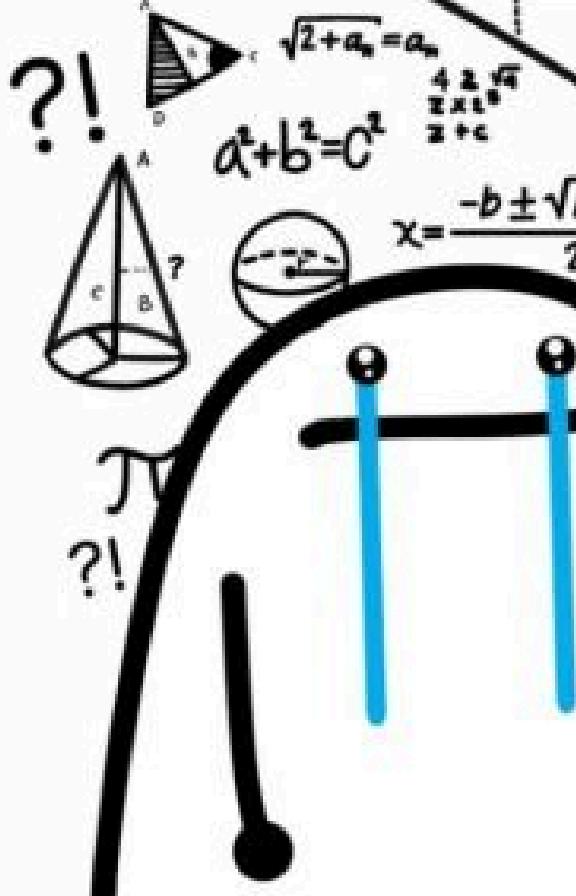


DESARROLLO DEL MODELO

¿QUÉ MODELOS SE PROBARON?

- SVM con kernel RBF → buen accuracy, pero no detectó clase minoritaria.
- Regresión Logística → interpretable pero con el mismo problema.
- Random Forest → logró identificar la clase minoritaria con métricas aceptables

| MODELO | ACCURACY | PRECISION (1) | RECALL (1) | F1-SCORE (1) | ROCAUC |
|---------------|----------|---------------|------------|--------------|--------|
| SVM | 0.957 | 0.000 | 0.000 | 0.000 | 0.685 |
| R. LOGÍSTICA | 0.957 | 0.000 | 0.000 | 0.000 | 0.363 |
| RANDOM FOREST | 0.968 | 0.500 | 0.333 | 0.400 | 0.861 |



EVALUACIÓN DEL MODELO

RANDOM FOREST, EL ELEGIDO

Accuracy

96.8%

Precision

50%

Recall

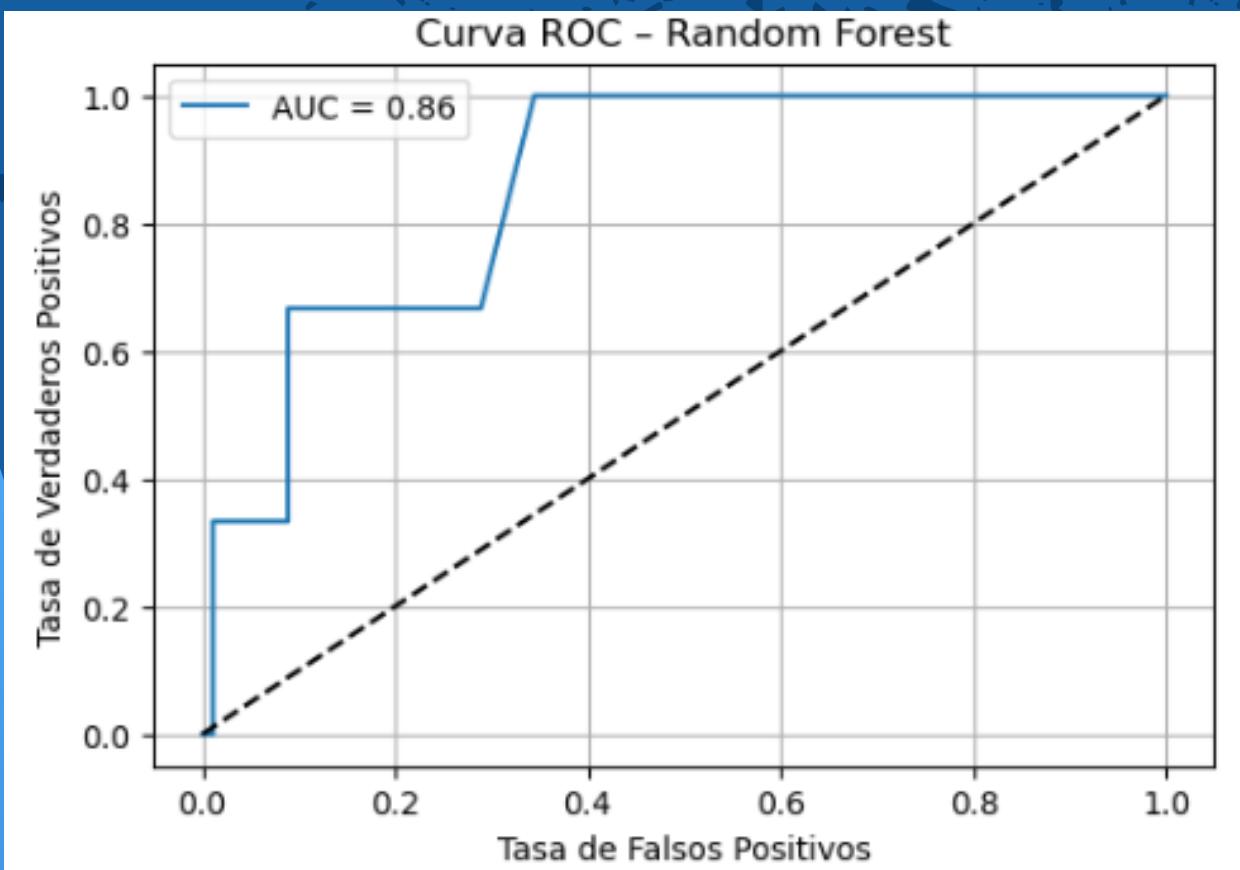
33%

F1-score

40%

ROC AUC

86.1%



- Alto accuracy → Pero NO es suficiente para evaluar la calidad del modelo.
- Precisión aceptable → Hubo varios falsos positivos.
- Baja Recall → Detectó algunos de los sismos que realmente fueron percibidos.
- F1-Score algo equilibrado → No logra un buen balance entre precisión y sensibilidad.
- ROC AUC → Separó muy bien las clases "percibido" y "no percibido".

CONCLUSIONES

¿QUÉ APRENDIMOS?

"El accuracy no lo es todo"

Aunque SVM y Regresión Logística mostraron alta exactitud, no detectaron ni un solo evento percibido. Esto demuestra que en datasets desbalanceados hay que mirar otras métricas.

"Detectar lo raro vale más que acertar lo común"

Random Forest fue el único que logró detectar parte de la clase minoritaria (sismos percibidos). En contextos como este, el valor está en detectar lo poco frecuente.

"Los datos bien tratados, predicen mejor"

La limpieza, estandarización y enriquecimiento del dataset fueron fundamentales para que el modelo pudiera aprender patrones útiles.

"La geografía sí importa"

Variables como la magnitud y la distancia mínima a ciudades fueron clave para predecir la percepción de un sismo. La cercanía importa.

RECOMENDACIONES Y TRABAJO FUTURO

¿CÓMO MEJORARLO?

- Incluir técnicas como SMOTE para balancear clases.
- Agregar nuevas variables: tipo de suelo, intensidad instrumental, duración.
- Integrar el modelo a sistemas en tiempo real para mejorar la comunicación preventiva.





GRACIAS!!!