

Entrenamiento del sistema

1.1 Datos

El modelo, que es la base fundamental de la aplicación, se usará para el cálculo de rutas seguras en España (nos limitamos a un país por ahora para validar el producto y luego expandirnos internacionalmente). Necesitamos, por tanto, datos de cualquier actividad criminal posible: asesinatos, robos, violaciones sexuales, tráfico de droga... Preferiblemente actualizados cada día y con una localización precisa (al menos el municipio del crimen).

Los datos que usaremos, por lo menos al inicio, serán datos del Ministerio de Interior (no tendríamos datos de reportes todavía), que nos proporcionan información de criminalidad agregada por municipios y trimestralmente (<https://estadisticasdecriminalidad.ses.mir.es/publico/portalestadistico/balances>).

La licencia de uso/reutilización de dichos datos tiene las siguientes condiciones (sacado de <https://estadisticasdecriminalidad.ses.mir.es/publico/portalestadistico/avisoLegal.html>):

Condiciones generales para la reutilización

Son de aplicación las siguientes condiciones generales para la reutilización de los documentos sometidos a ellas:

- Está prohibido desnaturalizar el sentido de la información.
- Debe citarse la fuente de los documentos objeto de la reutilización. Esta cita podrá realizarse de la siguiente manera: "Origen de los datos: Portal Estadístico de Criminalidad".
- Debe mencionarse la fecha de la última actualización de los documentos objeto de la reutilización, siempre cuando estuviera incluida en el documento original.
- No se podrá indicar, insinuar o sugerir que el Portal Estadístico de Criminalidad reutilizado participa, patrocina o apoya la reutilización que se lleve a cabo con él.
- Deben conservarse, no alterarse ni suprimirse los metadatos sobre la fecha de actualización y las condiciones de reutilización aplicables incluidos, en su caso, en el documento puesto a disposición para su reutilización.

Para lidiar con el problema de que los datos son publicados con una frecuencia baja (cada 3 meses), usaremos datos de noticias diarias, ya que no estamos utilizando la propiedad intelectual como el texto sino solo extraemos los hechos a partir del texto (<https://dudas.derechosdigitales.org/caso/como-pueden-usarse-las-noticias-los-articulos-y-las-fotos-de-prensa/>). Para no tener una única fuente de información, nos centraremos en 2 periódicos gratuitos sin muros de pago:

- eldiario.es (<https://www.eldiario.es/temas/asesinatos/>)
- 20minutos.es (<https://www.20minutos.es/tags/temas/asesinatos.html>)

A partir de sus avisos legales ([20minutos](#), [eldiario](#)), concluimos que hay que ponerse en contacto con dichos periódicos para conseguir la licencia de uso.

En caso de no conseguirlo, podemos usar servicios de agregadores de noticias <https://qnews.io/>, que proporcionan listado de noticias con encabezados y pequeñas descripciones para uso comercial con una suscripción de pago.

Esta segunda fuente (las noticias), al lanzar la aplicación se podría reemplazar o complementar con los reportes de los usuarios de la app.

1.2 Evaluación

En GPS-Safe evaluamos el modelo para asegurarnos de que, como sistema de alto riesgo, funcione de forma segura, fiable y ética antes de desplegarse en un entorno real. Vamos a definir la evaluación del modelo según los apartados especificados:

- a) ¿Cómo se evalúa que el modelo funcione?

La evaluación se basa en revisar cómo calcula el coeficiente de seguridad de cada zona, que es una suma con pesos para cada tipo de crimen (por ejemplo, $0.1 * \text{número de robos} + 0.2 * \text{número de violaciones} + \dots$). Para ello, combinamos pruebas de rendimiento y supervisión ética, en las que nos aseguramos de que el modelo pueda:

1. Integrar correctamente distintos tipos de datos (como criminalidad, reportes de usuario y contexto)
2. Predecir con alta precisión (el objetivo marcado es de un accuracy $> 95\%$)
3. Minimizar errores críticos. Especialmente, los falsos negativos, que podrían señalar como segura una zona que no lo es.
4. Ser explicable, cumpliendo los requisitos de interpretabilidad (XAI) para que sus decisiones puedan entenderse y revisarse.

- b) ¿Qué métricas se definen y usan para testearlo?

Se emplean métricas estándar para evaluar la calidad de las predicciones:

- MAE: mide cuánto se equivoca el modelo de media
- MSE: penaliza más los errores grandes
- F1: para pasar a esta métrica de “clasificación” pondremos un threshold para decidir si es muy peligroso o poco peligroso. Combina precisión y recall para evaluar el equilibrio del modelo.

- c) ¿Qué factores se están midiendo?

Los factores evaluados cubren rendimiento y comportamiento ético:

1. Precisión: que el modelo calcule bien el nivel de riesgo de cada zona
2. Generalización: que funcione bien con nuevos datos y se mantenga estable a largo plazo
3. Ausencia de sesgos (discriminaciones o patrones injustos). Esto será revisado por el Comité Ético.

4. Interpretabilidad (XAI): que sus decisiones puedan entenderse y justificarse y no actúe como una “caja negra”.
- d) ¿Hay algún benchmark existente que se pueda usar de referencia?

Sí. La evaluación aquí combina un objetivo interno y dos referencias científicas externas:

Objetivo interno: Como sistema de alto riesgo, GPS-Safe fija un listón propio muy alto:

- Precisión mínima: >95%
- Requisito crítico: cero falsos negativos aceptables

Referencias externas: Usamos dos estudios como guía metodológica:

1. Predicción de accidentes de tráfico (Random Forest + análisis espacial).

Explica el ≈78% de la varianza (buen rendimiento). Útil como referencia para combinar ML con análisis de red para rutas seguras.

Referencia: [Machine learning for predictions of road traffic accidents and spatial network analysis for safe routing on accident and congestion-prone road networks - ScienceDirect](https://www.sciencedirect.com/science/article/pii/S0960020316300302)

2. Identificación de rutas seguras según crimen (NLP + datos policiales).

≈89% de precisión en clasificación de noticias y ≈84% en identificación de ubicación; alta validación de usuarios (8'75/10). Relevante aquí porque muestra cómo mezclar fuentes de datos para estimar el riesgo urbano.

Referencia: [Analyzing Newspaper Crime Reports for Identification of Safe Transit Paths](https://www.sciencedirect.com/science/article/pii/S0960020316300302)

En conjunto, estos benchmarks demuestran que la metodología de GPS-Safe (IA + datos mixtos + análisis espacial) es viable, aunque nuestro objetivo interno (>95%) es más exigente que los resultados habituales en la literatura. Sin embargo, como es de cara a un Horizonte 2029, nos mostramos optimistas al respecto.

- e) Basadas en estas evaluaciones, benchmarks y testing, ¿qué objetivos debe cumplir el modelo para considerarlo adecuado para salir de la fase de testing a un entorno de producción en entorno real? (e.g. llegar a tal umbral “go/no-go” en tal evaluación)

Para que el modelo se considere adecuado para salir de la fase de testing a un entorno de producción en entorno real, debe cumplir cinco condiciones claras:

1. Umbral de precisión (“Go”): Debe superar el 95% de precisión en las pruebas.
2. Seguridad crítica (“No-Go”): No puede generar falsos negativos.

3. Cumplimiento ético y legal (“No-Go”): El modelo no puede ser una caja negra ni mostrar sesgos. Requiere validación del Comité Ético y cumplir los requisitos mínimos de interpretabilidad.
4. Estabilidad operativa (“No-Go”): No debe presentar fallos graves que puedan llevar a recomendar rutas con riesgo alto o conocido.
5. Supervisión continua (requisito permanente): Debe permitir monitorización y pruebas en funcionamiento para asegurar que el nivel de seguridad se mantiene durante su operación real.

En resumen, el modelo no solo debe alcanzar un alto rendimiento, sino demostrar que es seguro, interpretable y estable para poder pasar a producción.

Analogía sugerida por la IA: Pensar en la evaluación del modelo de GPS-Safe es como someter a prueba un nuevo modelo de airbag en un coche. No solo se mide qué tan preciso es el sensor al detectar un choque (la *accuracy* o la capacidad de procesar variables), sino que el umbral de “Go/No-Go” real es la **eliminación total de fallos de tipo falso negativo** (que el airbag no se despliegue cuando *realmente* se necesita). Además, el coche debe pasar una inspección de seguridad humana (Comité Ético) para asegurar que su diseño no perjudica a ningún conductor de manera discriminatoria (evitando sesgos) antes de que se le permita circular en carretera.

1.3 Hardware

Dado que el modelo GPS-Safe combina redes neuronales y procesamiento de redes espaciales sobre datos de criminalidad y tráfico, el hardware necesario se puede estimar de la siguiente manera:

Hardware recomendado

Se recomienda usar al menos una GPU **NVIDIA H100**.

Las H100 son GPUs de última generación, con hasta **989 TFLOPS** en operaciones tensoriales, lo que permite entrenar modelos complejos rápidamente.

Tiempo estimado

Para entrenar un ciclo completo (50 iteraciones de desarrollo) se estima **10 h de GPU**, considerando un uso intensivo y overhead de operaciones no optimizadas.

Se puede duplicar este tiempo por pruebas, debugging y ajustes durante el desarrollo (+100% margen).

Coste estimado en AWS:

- Instancia con 1 GPU H100 equivalente cuesta ~4,92 USD/h ($\approx 4,58 \text{ €/h}$)
- Por ciclo: $10 \text{ h} \times 4,58 \text{ €} \approx 45,8 \text{ €}$
- Para 50 iteraciones: $45,8 \times 50 \approx 2.290 \text{ €}$
- Con margen de desarrollo (+100%): hasta **4.580 €**

Coste eléctrico en España:

- Potencia GPU: $0,7 \text{ kW} \times 10 \text{ h} \approx 7 \text{ kWh}$ por ciclo

- Precio electricidad: 0,223 €/kWh → $7 \times 0,223 \approx 1,56 \text{ € por ciclo}$
- Para 50 iteraciones: $1,56 \times 50 \approx 78 \text{ €}$
- Con margen: hasta **156 €**

Huella de carbono:

- Energía consumida por ciclo: 7 kWh
- Intensidad de carbono electricidad España: 200 g CO₂/kWh
- Emisión: $7 \times 200 \text{ g} \approx 1,4 \text{ kg CO}_2 \text{ por ciclo}$
- Para 50 iteraciones: $1,4 \times 50 \approx 70 \text{ kg CO}_2$
- Con margen de desarrollo (+100%): hasta **140 kg CO₂**

Cifras anuales:

Nvidia H100
Power Draw:700W
Daily Cost:\$2.28
Monthly Cost:\$68.42
Annual Cost:\$832.42
Annual Carbon Cost:2215.18 lbs of CO ₂