



UA

**Universidad
de Alicante**

ESCUELA POLITÉCNICA SUPERIOR

Ejercicio de Retos

Curso académico 2024 / 2025

Semestre 2

Grado en Ingeniería en Inteligencia Artificial

Grupo 1

Profesora: Ricardo Moreno Rodríguez

Alumno: Santiago Álvarez Geanta

ÍNDICE

Bloque 1: Introducción y análisis del reto urbano	2
1.1. Contexto general	2
1.2. Justificación del reto elegido	2
1.3. Primeros dilemas técnicos y éticos	2
1.4. Primeras decisiones tomadas en este ejemplo	3
Bloque 2: Tipos de datos y decisiones sobre su obtención	4
2.1. ¿Qué datos necesita este sistema?	4
2.2. ¿De dónde se obtienen estos datos?	4
2.3. Alternativas de datos y decisiones	5
2.4. Riesgos asociados y medidas preventivas	5
Bloque 3: Elección del modelo de IA	6
3.1. ¿Qué tipo de modelo se necesita en este reto?	6
3.2. ¿Qué modelos puedo usar y por qué elegir uno u otro?	6
3.3. Comparación con un modelo más simple	7
Bloque 4: Selección y justificación de la infraestructura HPC	8
4.1. ¿Qué recursos de computación hacen falta?	8
4.2. Opciones de infraestructura posibles	9
4.3. Coste energético y sostenibilidad	9
4.4. Medidas para reducir el impacto ambiental	10
Bloque 5: Diseño del flujo de trabajo completo	11
5.1. ¿Cuánto consume entrenar este modelo?	11
5.2. ¿Qué pasa si se usa un modelo más sencillo?	11
5.3. ¿Vale la pena el coste energético?	12
5.4. Decisión final justificada	13
Bloque 6: Reflexión ética y técnica sobre el sistema propuesto	13
6.1. ¿Por qué es importante esta reflexión?	13
6.2. Reflexión sobre las decisiones técnicas	14
6.3. Reflexión sobre los valores implicados	15
6.4. ¿Qué harías distinto si el contexto cambiara?	15
Bloque 7: Comentario final y adaptaciones posibles	16
7.1. ¿Puede aplicarse este sistema a otros contextos?	16
7.2. ¿Qué se tendría que cambiar?	16

Bloque 1: Introducción y análisis del reto urbano

1.1. Contexto general

En el contexto de las ciudades inteligentes, anticipar la ocupación de espacios públicos como parques, zonas verdes o playas se ha convertido en un reto urbano clave, especialmente ante fenómenos como el cambio climático, el turismo masivo y la necesidad de garantizar el bienestar ciudadano. La afluencia a estos espacios varía según factores como la meteorología, eventos locales, días festivos, patrones de movilidad o incluso tendencias en redes sociales.

Un mal dimensionamiento o falta de previsión puede provocar aglomeraciones, degradación ambiental, falta de mantenimiento adecuado o situaciones de desigualdad en el acceso a estos espacios.

La Inteligencia Artificial, combinada con computación de alto rendimiento (HPC), permite analizar grandes volúmenes de datos en tiempo real para anticipar estos patrones de ocupación y facilitar una gestión más equitativa, sostenible y segura.

“He elegido el problema de la anticipación de la ocupación en espacios públicos porque permite aplicar modelos predictivos que ayudan a mejorar la sostenibilidad, el uso eficiente de recursos urbanos y la equidad ciudadana, apoyándose en infraestructuras HPC para procesar datos complejos y dinámicos.”

1.2. Justificación del reto elegido

Este reto resulta adecuado para aplicar técnicas de HPC e IA por varios motivos:

- Involucra grandes volúmenes de datos heterogéneos (sensores, climatología, redes sociales, datos de movilidad).
- Requiere modelos capaces de detectar patrones espacio-temporales complejos y de adaptarse a cambios frecuentes.
- Presenta implicaciones éticas relevantes sobre privacidad, acceso equitativo a servicios urbanos y vigilancia.
- Tiene un impacto directo en la calidad de vida, la sostenibilidad ambiental y la planificación urbana.
Además, se alinea con los contenidos de la asignatura sobre análisis de datos masivos, escalabilidad de sistemas IA y eficiencia energética.

1.3. Primeros dilemas técnicos y éticos

Desde el inicio surgen interrogantes importantes:

- ¿Qué fuentes de datos son fiables para estimar ocupación en tiempo real o predecirla con antelación?
- ¿Hasta qué punto podemos utilizar datos geolocalizados sin comprometer la privacidad?
- ¿Cómo evitar que zonas con menos dispositivos de monitoreo queden infrarrepresentadas?
- ¿Cómo equilibrar la precisión del modelo con su coste energético y su sostenibilidad?

Todo sistema urbano inteligente debe encontrar un equilibrio entre exactitud técnica, respeto a la privacidad, accesibilidad universal y eficiencia computacional. Estas preguntas iniciales son clave para orientar las decisiones de diseño.

1.4. Primeras decisiones tomadas en este ejemplo

En este caso, se ha decidido:

- Trabajar con una granularidad media: nivel de zonas urbanas o sectores costeros (no individual), para proteger la privacidad.
- Usar datos agregados provenientes de sensores de ocupación, cámaras anonimizadas, pronósticos meteorológicos y flujos de movilidad pública.
- Integrar información complementaria como eventos programados o aforos permitidos.
- Diseñar un sistema distribuido que permita adaptarse de forma continua a nuevos datos y situaciones no previstas (por ejemplo, olas de calor, restricciones por salud pública).

“Para este reto, he decidido trabajar a nivel de zonas urbanas y costeras, utilizando datos agregados de sensores, movilidad y condiciones ambientales, priorizando la protección de la privacidad y la adaptabilidad del sistema en distintos contextos urbanos.”

Bloque 2: Tipos de datos y decisiones sobre su obtención

2.1. ¿Qué datos necesita este sistema?

Para anticipar con precisión la ocupación de espacios públicos urbanos como parques o playas, el sistema debe alimentarse de datos que reflejen el comportamiento humano, las condiciones ambientales y la dinámica urbana.

Los tipos de datos considerados esenciales para este caso son:

- **Datos de ocupación histórica:** registros agregados del nivel de afluencia en zonas verdes y playas (por ejemplo, recuento de personas por franja horaria, sensores o imágenes anonimizadas).
- **Datos meteorológicos:** temperatura, humedad, velocidad del viento, radiación solar o precipitaciones, ya que influyen directamente en la decisión de las personas de acudir a espacios exteriores.
- **Datos de movilidad urbana:** desplazamientos desde zonas residenciales hacia áreas verdes, uso del transporte público hacia zonas costeras, etc.
- **Calendario de eventos y días festivos:** festividades locales, fines de semana, vacaciones escolares o eventos culturales que afectan significativamente la afluencia.

“Para poder hacer predicciones fiables en el problema de anticipación de ocupación en espacios públicos, necesito datos sobre afluencia histórica, condiciones meteorológicas y patrones de movilidad, porque influyen directamente en el comportamiento que quiero anticipar.”

2.2. ¿De dónde se obtienen estos datos?

La viabilidad del sistema depende de una obtención legal, ética y técnica de los datos. En este caso:

- **Datos de ocupación:** pueden obtenerse mediante sensores instalados por los municipios (por ejemplo, contadores de paso, cámaras con procesamiento de imagen anonimizada) o a través de plataformas Smart City ya existentes.
- **Datos meteorológicos:** accesibles mediante APIs públicas como AEMET, Meteocat o OpenWeather.
- **Datos de movilidad:** posibles a través de datos anonimizados de telefonía móvil (acuerdos con operadoras), redes de transporte público o plataformas como Google Mobility Reports.
- **Fechas especiales y eventos:** se pueden extraer de calendarios municipales abiertos, plataformas turísticas o incluso redes sociales con técnicas de análisis de tendencias.

En ningún caso se utilizarán datos con identificadores personales. Todos los datos deben estar previamente agregados o anonimizados, respetando la normativa vigente en protección de datos (como el RGPD).

2.3. Alternativas de datos y decisiones

Durante el diseño se han tomado decisiones para equilibrar precisión, sostenibilidad y ética:

- Se ha optado por trabajar en zonas piloto con alta densidad poblacional o turística (por ejemplo, playas urbanas o grandes parques metropolitanos).
- Los datos se agruparán por franja horaria (mañana, tarde, noche) en lugar de por hora exacta, para reducir el coste computacional y preservar la privacidad.
- Se incorporarán solo variables directamente relevantes para la predicción (clima, movilidad, eventos), dejando fuera aspectos más sensibles como perfil socioeconómico.

“He decidido limitar los datos a ocupación histórica agregada, clima, movilidad y eventos, para reducir el riesgo de sobreajuste y garantizar la privacidad. Aunque se pierde precisión horaria, se gana en sostenibilidad y replicabilidad.”

2.4. Riesgos asociados y medidas preventivas

Riesgo	Medida de prevención
Uso excesivo de datos personales	Agregación de datos por zona y franjas horarias
Representación desigual de zonas (sesgos urbanos)	Validación cruzada entre diferentes tipos de áreas (céntricas/periféricas)
Datos meteorológicos incompletos o inexactos	Contraste entre varias fuentes (AEMET + OpenWeather)
Dependencia excesiva de sensores físicos	Complemento con datos satelitales o estimaciones por IA

Estas medidas buscan anticipar los riesgos habituales en proyectos urbanos basados en datos, reforzando la confianza ciudadana y la fiabilidad del sistema.

Bloque 3: Elección del modelo de IA

3.1. ¿Qué tipo de modelo se necesita en este reto?

El problema que abordamos consiste en anticipar la ocupación futura de espacios públicos como playas y parques urbanos, basándonos en datos pasados y condiciones contextuales (como el clima, movilidad o eventos). Esto se considera un problema de **predicción de series temporales multivariadas**, ya que la variable objetivo, (el nivel de ocupación) varía con el tiempo y está influida por múltiples factores.

Ejemplos de patrones que el modelo debe ser capaz de aprender:

- Que las playas urbanas se llenan más los fines de semana soleados.
- Que los parques tienen más uso por la tarde entre semana.
- Que en días festivos o durante eventos culturales, algunas zonas se saturan.
- Que la ocupación baja si hay lluvia o viento fuerte, independientemente del día.

Además, la predicción debe ser sensible a contextos variables como vacaciones escolares, olas de calor o restricciones por salud pública.

“Este es un problema donde la variable principal cambia con el tiempo, así que necesito un modelo capaz de detectar patrones horarios, semanales y estacionales. Además, quiero que el modelo tenga en cuenta otras variables como el clima, la movilidad y los eventos locales.”

3.2. ¿Qué modelos puedo usar y por qué elegir uno u otro?

Hemos comparado varios modelos de predicción temporal en función de su idoneidad para manejar múltiples variables y su eficiencia computacional:

Modelo	Ventajas principales	Limitaciones o riesgos
ARIMA	Sencillo, interpretable, útil si solo se analiza una variable de ocupación	No maneja bien múltiples variables ni cambios bruscos
Random Forest	Robusto y rápido, útil para relaciones no lineales entre variables	No está diseñado para secuencias temporales directamente

XGBoost	Preciso en muchos entornos, eficiente con datos tabulares	Requiere ingeniería de características temporales manual
LSTM	Especializado en secuencias temporales, permite entrada multivariable, detecta ciclos	Entrenamiento más complejo y consumo computacional mayor

En nuestro caso, elegimos el modelo **LSTM (Long Short-Term Memory)**, un tipo de red neuronal recurrente, porque:

- Detecta **patrones temporales complejos** (por hora, día, estación).
- Permite incorporar **múltiples variables externas** como clima, movilidad o eventos.
- Se adapta bien a datos con **tendencias estacionales o repetitivas**.
- Puede ajustarse a diferentes escalas temporales (por ejemplo, predicción por hora o por día).

“He elegido LSTM porque es un modelo capaz de aprender patrones temporales y usar múltiples fuentes de información al mismo tiempo. Esto permite anticipar la ocupación de zonas públicas teniendo en cuenta tanto el historial como factores externos.”

3.3. Comparación con un modelo más simple

Para valorar esta elección, se ha comparado LSTM con **Random Forest**, un modelo más simple y ligero computacionalmente. Aunque Random Forest puede ofrecer resultados aceptables si se transforman las variables temporales (día, hora, clima), no tiene una estructura interna que comprenda secuencias, por lo que pierde capacidad de anticipación a largo plazo o en entornos muy dinámicos.

Conclusión comparativa:

- LSTM ofrece mejor rendimiento si se dispone de datos temporales suficientemente largos y variados.
- Random Forest puede ser útil en entornos con menos recursos computacionales o en zonas donde solo se necesiten predicciones de corto plazo.

Bloque 4: Selección y justificación de la infraestructura HPC

4.1. ¿Qué recursos de computación hacen falta?

Una vez elegido el modelo LSTM para anticipar la ocupación de zonas públicas, es necesario evaluar los recursos computacionales requeridos para su entrenamiento.

Este modelo necesita procesar grandes volúmenes de datos históricos, incluyendo variables climáticas, de movilidad y eventos, segmentadas por zona y momento del día. Entrenar un LSTM con múltiples entradas y secuencias largas requiere una **capacidad de cálculo intensiva**, especialmente en operaciones matriciales.

Esto implica que:

- No es viable entrenar el modelo de forma eficiente en un portátil personal.
- Se requiere **hardware con aceleración por GPU** y capacidad de almacenamiento para gestionar datasets multivariantes y de larga duración temporal.

“He decidido entrenar el modelo en un entorno con acceso a GPU, ya que este tipo de red neuronal necesita operaciones matriciales intensivas que se aceleran con este hardware.”

4.2. Opciones de infraestructura posibles

Opción	Ventajas	Inconvenientes
Clúster local (universidad)	Control total sobre los datos, privacidad asegurada	Limitaciones de acceso, mantenimiento, y disponibilidad
Servidor con GPU dedicada	Rendimiento ideal para redes LSTM	Alto coste económico, posible infrautilización
Plataforma cloud HPC	Escalabilidad, pago por uso, rápida configuración	Dependencia de terceros, impacto ambiental si no optimizada
Entrenamiento distribuido	Permite acelerar el proceso usando varios nodos	Requiere conocimientos técnicos en paralelización y gestión

En este caso, **optamos por una plataforma en la nube con GPU**, para facilitar el escalado del modelo si se quiere aplicar a distintas zonas o ciudades, y permitir reentrenamientos periódicos bajo demanda.

4.3. Coste energético y sostenibilidad

Entrenar modelos de IA, especialmente redes neuronales profundas, implica un **consumo energético no trivial**. Por ejemplo:

- Entrenar un modelo LSTM básico puede requerir unas 4 horas en una GPU estándar (~300W), lo que equivale a **1,2 kWh por sesión de entrenamiento**.
- Si se realizan 10 sesiones para ajustar hiperparámetros y validar el modelo, se alcanzarían **12 kWh**, aproximadamente igual al consumo de un frigorífico eficiente durante 2 días.

Ahora bien, si el sistema se amplía a múltiples zonas de una ciudad (ej. 20 parques + 10 playas), y se actualiza cada semana, el impacto energético escalaría rápidamente.

“A la hora de decidir el modelo y su infraestructura, debemos valorar no solo la precisión técnica, sino también su coste energético. Un modelo menos complejo puede ser preferible si el sistema va a operar de forma continua y a gran escala.”

4.4. Medidas para reducir el impacto ambiental

Para mitigar el impacto ecológico, proponemos las siguientes medidas:

- **Entrenar un único modelo general para la ciudad** (o pocos modelos regionales), en lugar de uno por espacio público.
- Ejecutar entrenamientos en **horarios valle** para reducir demanda energética y costes.
- Activar el **reentrenamiento automático solo cuando haya desviaciones significativas** en los patrones de uso detectados.
- **Documentar el consumo energético estimado** en cada ciclo de entrenamiento para facilitar futuras decisiones de optimización.

“He optado por reducir el número de reentrenamientos para minimizar el impacto ambiental, y solo actualizo el modelo cuando los datos muestran un cambio significativo en los patrones de ocupación.”

Bloque 5: Diseño del flujo de trabajo completo

5.1. ¿Cuánto consume entrenar este modelo?

El modelo LSTM seleccionado para anticipar la ocupación en espacios públicos necesita un entrenamiento exigente, especialmente al considerar múltiples años de datos y variables contextuales como clima, eventos y patrones de movilidad.

Supuestos para estimar el consumo energético:

- Entrenamiento de 4 horas por sesión en una GPU de 300 W
- 10 sesiones necesarias para ajustar y validar el modelo
- 10 zonas diferentes analizadas (playas, parques, plazas...)

Cálculo total:

- $4 \text{ h} \times 300 \text{ W} = 1,2 \text{ kWh por sesión}$
- $1,2 \text{ kWh} \times 10 \text{ sesiones} \times 10 \text{ zonas} = 120 \text{ kWh}$

Interpretación:

Ese consumo es similar al gasto mensual de un hogar eficiente o al de cargar un vehículo eléctrico para recorrer unos 600 km. Aunque puede parecer razonable en un entorno controlado o académico, en despliegues municipales a gran escala este consumo puede multiplicarse considerablemente.

5.2. ¿Qué pasa si se usa un modelo más sencillo?

Modelos estadísticos clásicos como **ARIMA** podrían ser suficientes si los patrones de ocupación son relativamente regulares y si se prioriza la eficiencia sobre la precisión.

Ventajas de ARIMA en este contexto:

- Bajo consumo energético (puede entrenarse en un portátil estándar).
- Explicable: útil si se necesita justificar el funcionamiento ante entidades públicas o ciudadanos.
- Aceptable cuando los patrones de uso son previsibles (como fines de semana o temporadas turísticas).

Estimación de consumo:

- 0,1 kWh por entrenamiento \times 10 sesiones \times 10 zonas = **1 kWh**

Comparativa rápida:

- LSTM: 120 kWh
- ARIMA: 1 kWh
→ El modelo neuronal consume aproximadamente 120 veces más energía.

“Una alternativa más ligera como ARIMA permitiría ahorrar una gran cantidad de energía. Aunque se perdería precisión ante patrones menos regulares, el modelo sería más sostenible y comprensible para su uso en contextos municipales.”

5.3. ¿Vale la pena el coste energético?

Modelo	Precisión esperada	Adaptabilidad	Consumo energético	Facilidad de explicar
ARIMA	Media	Limitada	Muy bajo	Alta
LSTM	Alta	Alta	Elevado	Baja

Evaluación del contexto:

- En playas o parques urbanos con mucha variabilidad en la ocupación (por clima, turismo, festivales, etc.), LSTM se adapta mejor.
- En zonas con patrones muy estables o para prototipos rápidos, puede bastar un modelo como ARIMA.

5.4. Decisión final justificada

En este proyecto, se ha optado por **LSTM** porque:

- La ocupación depende de **muchos factores externos** (clima, movilidad, calendario, eventos).
- Se buscan **patrones no evidentes** que cambian según el contexto.
- Es útil demostrar cómo **la IA puede integrarse en la planificación urbana moderna**.

“He elegido LSTM por su capacidad para trabajar con múltiples variables y aprender patrones complejos que afectan la ocupación en espacios públicos. Aun así, si el objetivo fuera reducir el consumo o escalar de forma sostenible, consideraría modelos más simples como ARIMA.”

Bloque 6: Reflexión ética y técnica sobre el sistema propuesto

6.1. ¿Por qué es importante esta reflexión?

Diseñar un sistema de IA no es solo un reto técnico: implica tomar decisiones que tienen impacto en las personas, en los espacios públicos y en el medio ambiente. Por eso es fundamental detenerse al final del diseño para pensar:

- ¿Qué elecciones hemos hecho en el modelo y la infraestructura?
- ¿Qué valores hemos priorizado (precisión, sostenibilidad, privacidad...)?
- ¿Qué limitaciones tiene nuestra propuesta?
- ¿Qué cambiaríamos si el entorno fuese distinto?

“No basta con que un modelo funcione. También debe ser justificable, comprensible, eficiente y respetuoso con su entorno.”

6.2. Reflexión sobre las decisiones técnicas

Modelo elegido: LSTM

- Se ha seleccionado LSTM por su capacidad para captar patrones de ocupación que cambian con el tiempo y dependen de muchas variables: clima, eventos, vacaciones, movilidad...
- Puede manejar datos complejos y no lineales, algo necesario para anticipar la presencia de personas en espacios abiertos.
- Requiere muchos datos históricos y una infraestructura de cómputo potente (GPU), lo que implica más consumo energético.
- Es menos interpretable que otros modelos: puede ser difícil explicar a técnicos municipales o ciudadanos cómo llega a una predicción.

Modelo alternativo: ARIMA u otro estadístico simple

- Fácil de implementar y entender.
- Consumo energético bajo, se puede entrenar en ordenadores normales.
- Funciona bien si los patrones son muy regulares, como la afluencia los fines de semana.
- No maneja bien múltiples variables ni situaciones imprevistas (lluvias súbitas, conciertos, alertas...).

Infraestructura HPC (con GPU):

- Necesaria para entrenar eficientemente un modelo LSTM para varias zonas urbanas.
- Tiene un impacto energético significativo si no se optimiza.
- Se propusieron medidas para minimizarlo: entrenar modelos generales, reducir frecuencia de actualización y usar horarios valle.

“Aunque he elegido un modelo que requiere más recursos, considero que sus ventajas en precisión y adaptabilidad lo justifican en este caso. Sin embargo, reconozco que en otros contextos, una solución más simple habría sido más adecuada.”

6.3. Reflexión sobre los valores implicados

Cada decisión refleja una prioridad. Este proyecto ha tratado de equilibrar diferentes valores:

Decisión tomada	Valor priorizado
Usar datos agregados por hora y zona	Protección de la privacidad
Elegir un modelo capaz de integrar muchas variables	Precisión y realismo en la predicción
Comparar con modelos más simples	Responsabilidad técnica y transparencia
Proponer medidas para reducir el consumo	Sostenibilidad y uso eficiente de los recursos
Diseñar un modelo único para toda la ciudad	Escalabilidad y eficiencia computacional

“He intentado equilibrar precisión técnica con sostenibilidad y privacidad, eligiendo decisiones que me parecen justificables desde un punto de vista ético y realista.”

6.4. ¿Qué harías distinto si el contexto cambiara?

El sistema debe poder adaptarse a distintos escenarios. Algunas alternativas que se considerarían si cambian las condiciones:

- **Entorno con recursos limitados:** Se podría usar un modelo más ligero, como un árbol de decisión o ARIMA, para mantener bajo el consumo energético.
- **Falta de datos fiables:** El sistema podría basarse más en reglas simples (por ejemplo, ocupación media por día de la semana) en vez de aprendizaje automático.
- **Requisito de alta transparencia:** Se priorizarían modelos fácilmente explicables, aunque se perdiera algo de precisión.
- **Ámbitos rurales o con baja variabilidad:** Bastaría una solución más sencilla, incluso con visualización directa de datos históricos.

“Si el entorno de aplicación cambiara, ajustaría mi propuesta para mantener el equilibrio entre precisión, coste y sostenibilidad.”

Bloque 7: Comentario final y adaptaciones posibles

7.1. ¿Puede aplicarse este sistema a otros contextos?

Aunque este sistema se ha diseñado para anticipar la ocupación en zonas verdes y playas, su lógica puede adaptarse fácilmente a otros espacios o necesidades. La clave está en el enfoque: analizar patrones de comportamiento colectivo a partir de datos temporales y contextuales.

Posibles adaptaciones:

- **Transporte público:** Predecir la ocupación de estaciones o líneas de metro/bus en horas punta.
- **Eventos o instalaciones deportivas:** Anticipar afluencia a estadios, gimnasios o centros deportivos municipales.
- **Centros de salud o atención ciudadana:** Prever momentos de mayor afluencia y mejorar la organización de turnos o recursos.
- **Turismo en entornos rurales o de costa:** Predecir flujos de visitantes en parques naturales, senderos o zonas costeras más alejadas.
- **Entornos educativos:** Anticipar la ocupación en bibliotecas o comedores escolares en función del calendario académico o meteorología.

7.2. ¿Qué se tendría que cambiar?

Para adaptar el sistema a otro escenario, habría que revisar:

- **Variables externas relevantes:** Por ejemplo, en transporte importa la hora punta laboral, en playas el tiempo atmosférico.
- **Frecuencia y tipo de datos:** En centros de salud pueden ser diarios; en playas, cada 15 minutos en verano.
- **Modelo más adecuado:** Puede que en otros casos un modelo más simple sea suficiente.
- **Objetivos del sistema:** En algunos contextos no interesa tanto predecir personas, sino flujos, colas o tiempos de espera.