
This is the **published version** of the bachelor thesis:

Merroun Merroun, Adam; Castelló Bueno, Marc, dir. Monitoratge de la Qualitat de l'Aire amb LoRaWAN i Machine Learning. 2025. 9 pag. (Grau en Gestió de Ciutats Intel·ligents i Sostenibles)

This version is available at <https://ddd.uab.cat/record/317291>

under the terms of the  license

Monitoratge de la Qualitat de l'Aire amb LoRaWAN i Machine Learning

Adam Merroun Merroun

Resum— Els models de machine learning han transformat nombrosos àmbits, destacant especialment en la presa de decisions automatitzada. Aquests models permeten analitzar grans volums de dades en temps real, identificar patrons complexos i optimitzar processos, millorant així l'eficiència i la capacitat de resposta en diverses situacions. Aquest projecte té com a objectiu combinar les tecnologies LoRaWAN i Machine Learning per aconseguir espais més intel·ligents i automatitzats.

Paraules clau— Internet of Things, LoRaWAN, Machine Learning.

Abstract— Machine learning models have transformed numerous fields, particularly in automated decision-making. These models allow for the analysis of large volumes of data in real-time, identifying complex patterns and optimizing processes, thus improving efficiency and responsiveness in various situations. This project aims to combine LoRaWAN technology and Machine Learning to create smarter and more automated spaces.

Keywords— Internet of Things, LoRaWAN, Machine Learning.

1 INTRODUCCIÓ - CONTEXT DEL TREBALL

ACTUALMENT, els problemes ambientals tenen una repercussió creixent en la salut i el benestar de les persones. Diversos estudis han demostrat que una mala qualitat de l'aire en espais interiors pot generar impactes com fatiga, mals de cap o pèrdua de concentració, especialment en entorns educatius on es requereix un alt nivell de rendiment cognitiu. Per exemple, segons Satish et al. (2012), concentracions de CO₂ superiors a 1000 ppm poden afectar directament la presa de decisions i l'eficiència mental.

Les tecnologies actuals permeten abordar aquests reptes de manera més efectiva. Les xarxes IoT (Internet de les Coses), com LoRaWAN, faciliten la transmissió remota de dades ambientals amb baix consum energètic i gran abast. Això les fa especialment adequades per a espais grans com universitats. Alhora, l'ús de models de Machine Learning, com l'algorisme Isolation Forest, permet analitzar aquestes dades en temps real i detectar anomalies ambientals sense necessitat d'intervenció humana.

- E-mail de contacte: adam.merroun@autonoma.cat
- Treball tutoritzat per: Marc Castelló Bueno (Departament de Geografia)
- Curs 2024/25

En aquest projecte, es proposa la combinació d'aquests dos enfocaments. A través de sensors com els Milesight AM107, s'obtenen dades com la temperatura, humitat i concentració de CO₂, que es processen per identificar situacions de risc ambientals dins de la universitat. Aquesta anàlisi pot servir com a base per automatitzar sistemes de ventilació o enviar alertes, millorant la qualitat de l'aire i la comoditat en els espais d'aprenentatge.

A més, aquesta proposta té un valor afegit: fomenta la participació de la comunitat universitària en l'ús de dades reals per afrontar problemàtiques ambientals. També pot ser una eina educativa per a estudiants interessats en camps com la ciència de dades, l'enginyeria ambiental i la gestió intel·ligent d'edificis.

2 OBJECTIUS

2.1 Objectiu general

Desenvolupar un model de detecció d'anomalies ambientals mitjançant tècniques de *machine learning*, entrenat amb dades reals proporcionades per sensors LoRaWAN instal·lats al campus universitari, amb l'objectiu d'avaluar el seu potencial en futurs sistemes intel·ligents de monitoratge de la qualitat ambiental.

2.2 Objectius específics

- Obtenir i analitzar dades ambientals (CO₂, temperatura, humitat) recollides prèviament per sensors LoRaWAN, validant-ne la qualitat mitjançant processos de neteja, transformació i preparació per a l'entrenament de models.
- Aplicar el model no supervisat *Isolation Forest* per detectar anomalies en els valors ambientals, identificant patrons atípics que podrien indicar situacions de risc en espais interiors.
- Interpretar els resultats del model per extreure conclusions sobre la seva fiabilitat i coherència amb estudis científics previs sobre qualitat ambiental en entorns educatius.
- Proposar millores en la infraestructura de monitoratge ambiental del campus, com l'ampliació de la xarxa de sensors i la integració futura amb actuadors automatitzats.
- Aportar una primera estimació dels costos i beneficis d'implementar aquest tipus de sistemes intel·ligents en entorns universitaris.

2.3 Hipòtesi

L'aplicació del model *Isolation Forest* a dades ambientals reals recollides mitjançant sensors LoRaWAN permet detectar anomalies de manera fiable i coherent amb els llinars crítics establerts per la literatura científica, fet que el converteix en una eina potencialment útil per a la millora de la gestió ambiental en espais universitaris.

2.4 Preguntes de recerca

- El model *Isolation Forest* és capaç de detectar de forma autònoma anomalies ambientals significatives a partir de dades reals obtingudes amb LoRaWAN?
- Les anomalies identificades (per exemple, CO₂ > 1000 ppm, T > 25°C, HR < 40% o > 60%) són consistents amb els criteris establerts per la recerca en qualitat ambiental?
- Quines aplicacions pràctiques es poden derivar d'aquest model en un entorn com la UAB, i com podrien contribuir al benestar i rendiment dels usuaris?
- Quins recursos tècnics i econòmics caldrien per implementar un sistema de monitoratge intel·ligent basat en aquest model a escala de campus?

3 MOTIVACIÓ

Durant la meua trajectòria com a estudiant de la UAB, he desenvolupat un interès creixent per l'aplicació de tecnologies intel·ligents en l'àmbit ambiental, especialment en la sensorització i l'anàlisi de dades mitjançant tècniques d'aprenentatge automàtic. Aquest interès s'ha vist reforçat en observar les condicions ambientals en diversos espais tancats del campus, com les aules, en què sovint es percep una qualitat de l'aire deficient.

Tal com apunten Satish et al. (2012), concentracions de CO₂ superiors a 1000 ppm poden afectar negativament la presa de decisions, mentre que Wargocki i Wyon (2006) demostren que temperatures per sobre dels 25°C redueixen el rendiment cognitiu. A més, Arundel et al. (1986) evidencien que una humitat relativa fora del rang òptim (entre el 40% i el 60%) pot afavorir la proliferació de microorganismes i incidir negativament en la salut. Aquests estudis posen de manifest que paràmetres ambientals sovint invisibles poden tenir efectes directes sobre el benestar i la productivitat en entorns educatius.

Tanmateix, el monitoratge contínuu d'aquests factors encara és poc habitual en espais universitaris. Aquesta realitat m'ha portat a voler explorar, a través d'aquest treball, com la combinació de xarxes de sensors IoT i models de *machine learning* pot facilitar la detecció d'anomalies ambientals de forma autònoma i eficient.

Aquest projecte representa una oportunitat per aplicar coneixements tècnics a una problemàtica real i concreta, alineant-se amb els objectius de sostenibilitat i qualitat de vida que una universitat com la UAB hauria de promoure. Crec fermament que la integració de sistemes intel·ligents pot contribuir a millorar l'entorn d'aprenentatge, anticipant-se a situacions adverses i promovent espais més sans, segurs i eficients.

4 ESTUDIS

-Segons Satish et al. (2012), exposar-se a concentracions de CO₂ superiors a 1000 ppm pot afectar negativament la capacitat de presa de decisions. L'estudi va demostrar que, a partir d'aquest llindar, es redueixen les nostres competències cognitives, com l'ús eficient de la informació, i aquesta s'accentua a mesura que puja la quantitat de ppm. Aquest estudi demostra que nivells moderats de CO₂, habituals en espais interiors mal ventilats, poden tenir un impacte directe sobre el rendiment cognitiu.

-Segons Wargocki i Wyon (2006), temperatures superiors als 25 graus poden provocar una disminució significativa del rendiment escolar i de les tasques de concentració. L'estudi demostra que una reducció de la temperatura ambient millora l'eficiència en la realització de tasques i la velocitat de processament mental. Així, mantenir una temperatura adequada és clau per assegurar un entorn d'aprenentatge òptim.

-Segons Arundel et al. (1986), nivells de humitat relativa per sota del 40 per cent o per sobre del 60 per cent poden tenir efectes adversos sobre la salut i la comoditat, ja que, una humitat inadequada pot afavorir la proliferació de bacteris, virus i fongs, a més d'afectar la percepció tèrmica i augmentar la sequedat de les mucoses. L'estudi defensa que mantenir la humitat relativa dins del rang òptim (40-60 per cent) és essencial per al benestar i la salut en espais interiors.

5 DEFINICIONS PREVIES

Aquest apartat introdueix els conceptes fonamentals per comprendre la metodologia que es presentarà més endavant, establint una base de coneixements que facilitaran l'aplicació dels processos i tècniques en el projecte.

5.1 Sensors LoRaWAN

És un tipus de xarxa de comunicació sense fils dissenyada per a dispositius IoT (Internet of Things). Aquesta tecnologia permet la transmissió de dades a llargues distàncies amb un consum d'energia molt baix, cosa que la fa ideal per a sensors ambientals en espais grans o amb difícil accés.

Els sensors utilitzats en aquest projecte mesuren paràmetres ambientals i envien la informació a través d'una passarel·la (gateway) cap a un servidor central, on les dades són processades i analitzades.

5.2 Model de Machine Learning

És una branca de la intel·ligència artificial que permet als sistemes identificar patrons i detectar desviacions sense supervisió humana constant. En aquest projecte, s'ha aplicat Isolation Forest, un algorisme que separa valors atípics en paràmetres com els nivells de CO₂ per prevenir problemes de qualitat de l'aire.

5.3 Variables ambientals

Temperatura (°C): Mesura de la calor o fred en un espai. Nivells inadequats poden afectar el confort i el rendiment cognitiu.

Humitat relativa (%): Proporció de vapor d'aigua en l'aire. Un entorn massa sec o humit pot influir en la salut i la qualitat de l'aire interior.

Diòxid de carboni (CO₂, ppm): Gas generat per la respiració humana. Valors superiors a 1.000 ppm poden afectar la concentració i el benestar.

6 METODOLOGIA

Per desenvolupar aquest estudi, s'ha dut a terme primer una anàlisi teòrica per comprendre el funcionament de les tecnologies implicades, amb especial atenció a la comunicació LoRaWAN i als models d'aprenentatge automàtic, concretament el model *Isolation Forest*. Aquesta fase inicial ha permès establir una base sòlida per entendre com aquestes eines poden aplicar-se al monitoratge ambiental.

L'estudi s'ha basat en dades ambientals reals proporcionades per sensors LoRaWAN instal·lats en diferents espais del campus de la UAB. Aquestes dades inclouen variables com la concentració de CO₂, la temperatura i la humitat relativa. L'objectiu ha estat analitzar aquestes dades i aplicar tècniques de *machine learning* per detectar anomalies ambientals que podrien indicar situacions de risc, amb la voluntat de formular propostes de millora.

Un cop obtingudes les dades, s'ha realitzat un procés de neteja i preprocessament per garantir-ne la fiabilitat: s'han eliminat valors erronis o incomplets, i s'han normalitzat les variables per tal de facilitar-ne l'anàlisi posterior. Aquestes transformacions són fonamentals per assegurar el correcte funcionament dels models d'aprenentatge automàtic.

Per a la detecció d'anomalies s'ha emprat el model *Isolation Forest*, especialment indicat per identificar valors atípics dins de conjunts de dades no etiquetades. Aquest model construeix arbres de decisió de manera aleatòria i identifica com a anòmals els valors que poden ser aïllats amb poques divisions. En aquest cas, l'algorisme s'ha entrenat principalment amb dades del paràmetre CO₂, tot i que és escalable a més dimensions.

Per validar el rendiment del model, s'ha utilitzat la tècnica habitual de dividir les dades en dos conjunts: el 80% per entrenament i el 20% per validació. Això ha permès avaluar la seva capacitat de detecció en condicions realistes. Els llindars considerats com a indicadors de risc han estat: CO₂ > 1000 ppm, temperatura > 25°C, i humitat relativa < 40% o > 60%, segons referències científiques.

Finalment, es proposa una estratègia d'ampliació de la xarxa de sensors, tant dins la UAB com en altres entorns, per aconseguir una cobertura més extensa i una millor capacitat de monitoratge. Aquesta ampliació hauria d'anar acompanyada d'un sistema d'integració amb actuadors intel·ligents per activar mesures correctores, com ara la ventilació automàtica, en cas de detectar-se anomalies significatives.

7 MODEL DE MACHINE LEARNING

7.1 Introducció al model de detecció d'anomalies

Per a moltes persones sense formació prèvia en aprenentatge automàtic, pot resultar complex entendre com un sistema pot detectar patrons estranys sense instruccions explícites. En aquest projecte, s'ha utilitzat un model anomenat *Isolation Forest*, que és especialment útil per detectar valors anòmals en conjunts de dades ambientals, com el CO₂, la temperatura o la humitat.

El funcionament d'aquest model es basa en la idea d'aïllar les dades que es comporten de manera diferent respecte a la resta. Aquestes dades "estranys" poden ser separades més ràpidament mitjançant una sèrie de divisions successives, com si es tractés de preguntes sí/no que permeten aïllar un valor del conjunt general.

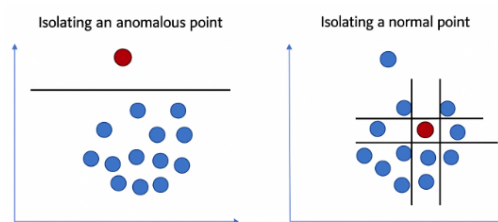


Fig. 1: Estructura bàsica de funcionament de l'algorisme Isolation Forest utilitzat per detectar anomalies en dades ambientals.

Aquest model és de tipus no supervisat, la qual cosa significa que no necessita dades etiquetades (és a dir, no cal indicar-li quins valors són “bons” o “dolents” prèviament). A partir de l'estructura dels valors recollits, és capaç de determinar quins es comporten de manera diferent.

Aquesta capacitat resulta especialment útil en escenaris on no sempre és evident què representa una situació de risc. Per exemple, un valor de CO₂ superior a 1000 ppm pot ser acceptable durant un interval curt de temps, però si es manté elevat persistentment pot ser un indicador d'un problema de ventilació greu. El model permet detectar aquests comportaments sense necessitat de definir manualment casos específics.

7.2 Càrrega i Preparació de Dades

Per al desenvolupament d'aquest model, es va utilitzar un conjunt de dades que prové d'un fitxer CSV amb mesuraments ambientals, incloent valors com el CO₂, temperatura i humitat. Aquest fitxer conté una gran quantitat de dades, però no tot estava en un format adequat per ser analitzat. Per això, es va dur a terme una sèrie de passos per a la neteja i preparació de les dades:

- Es va carregar el conjunt de dades en un DataFrame utilitzant la llibreria pandas que elimina la columna Unnamed: 0 (una columna no necessària generada per l'índex de l'arxiu CSV).

- Es va canviar els noms de les columnes per fer-les més comprensibles es va decidir eliminar la columna resultat, ja que no aportava informació rellevant per al model.

- També es va seleccionar les dades que contenen mesuraments que m'interessen com en aquest cas el CO₂, on es va eliminar les files amb valors nuls per evitar que afectessin la qualitat de les prediccions del model.

- Finalment es va aconseguir que la columna temps estigués en el format de data i hora (datetime) i es va convertir els valors de la columna valor a numèrics. Això va ser necessari per poder treballar amb els valors de manera coherent i adequada.

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd

arxiu_csv = 'data_feb_2025.csv'
df = pd.read_csv(arxiu_csv, comments='#')

if 'Unnamed: 0' in df.columns:
    df = df.drop(columns=['Unnamed: 0'])

df.columns = ['resultat', 'taula', 'inici', 'fi', 'temps', 'valor', 'camp', 'mesura', 'ubicacio']
df = df.drop(columns=['resultat'])

df_filtat = df[df['camp'].isin(['co2', 'temperatura', 'humiditat'])]
df_filtat = df_filtat.dropna(subset=['temps', 'inici', 'fi'])
df_filtat['valor'] = pd.to_numeric(df_filtat['valor'], errors='coerce')
df_filtat = df_filtat.dropna(subset=['valor'])
df_filtat['temps'] = pd.to_datetime(df_filtat['temps'], format='mixed', errors='coerce')
df_pivot = df_filtat.pivot(index='temps', columns='camp', values='valor')
df_pivot.reset_index(inplace=True)
print(df_pivot.head())
```

Fig. 2: Fragment de codi Python utilitzat per carregar i netejar les dades ambientals, incloent eliminació de columnes innecessàries.

camp	temps	ubicacio	co2	humidity
0	2023-07-25 13:48:55+00:00	eui-24e124128c147204	503.0	48.5
1	2023-07-25 13:50:42+00:00	am3019-testqc2090	472.0	36.0
2	2023-07-25 13:52:08+00:00	eui-24e124725c461468	461.0	36.5
3	2023-07-25 13:54:00+00:00	am307-9074	431.0	37.0
4	2023-07-25 13:54:19+00:00	eui-24e124128c147499	674.0	58.0

camp	temperature
0	25.6
1	27.8
2	27.5
3	27.4
4	24.8

Fig. 3: Dades correctament processades

7.3 Anàlisi i Visualització de les Dades

Un cop les dades van ser netejades i preparades, es va voler obtenir una comprensió general del comportament del CO₂ al llarg del temps. Per aquest motiu, es va realitzar una anàlisi de les mitjanes mensuals de cada variable:

Es van transformar les dates a un format mensual utilitzant una funció, de manera que cada registre representés el mes i l'any corresponent. Aquesta transformació va permetre observar les mitjanes mensuals per a cada mesura.

```
df_pivot['temps'] = pd.to_datetime(df_pivot['temps']).dt.tz_localize(None)
df_pivot['any_mes'] = df_pivot['temps'].dt.to_period('M')
df_mitjanes = df_pivot.groupby('any_mes')['co2'].mean()

fig, ax = plt.subplots(figsize=(10, 6))

# Gràfica CO2
ax.plot(df_mitjanes.index.astype(str), df_mitjanes, label='CO2', color='red', linestyle='-', marker='o')
ax.set_title('Mitjana mensual del CO2')
ax.set_xlabel('CO2 (ppm)')
ax.set_ylabel('Any-Mes')
ax.grid(True)
ax.legend()

plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

Fig. 4: Codi per anàlisi i visualització de les dades

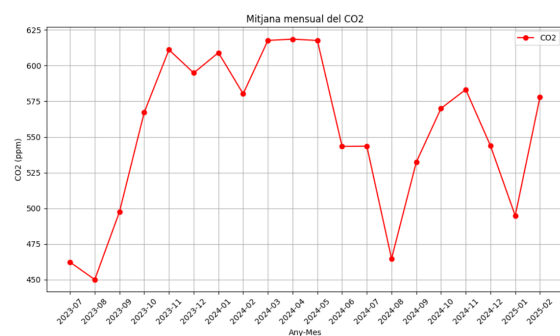


Fig. 5: Gràfica de mitjanes mensuals dels nivells de CO₂, obtinguda després de transformar les dates al format mes-any.

7.4 Preparació del model

Per preparar les dades per al model Isolation Forest, es va aplicar un escalat utilitzant la tècnica StandardScaler de Scikit-learn, assegurant que les dades de CO₂ tinguessin mitjana zero i desviació estàndard unitària. Aquest pas va garantir que tots els valors tinguessin la mateixa importància en el procés d'aprenentatge.

L'objectiu del model Isolation Forest és identificar anomalies, com ara pics inesperats en les concentracions de CO₂. Aquesta tècnica d'aprenentatge automàtic no supervisada crea diversos arbres d'aïllament per separar les dades, identificant els punts més fàcilment separables.

Es va configurar el model amb un paràmetre de contaminació de 0.1, indicant que aproximadament un 10 per cent de les dades serien anomalies. El model es va entrenar amb les dades escalades, permetent-li aprendre la distribució normal de CO₂ i identificar les anomalies.

```
df_co2 = df_pivot[['temps', 'ubicacio', 'co2']]

from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
df_co2_scaled = scaler.fit_transform(df_co2[['co2']])

# Ordenar les dades per temps
df_co2 = df_co2.sort_values(by='temps')

# Determinar el punt de divisió
punt_divisio = int(len(df_co2) * 0.8) # Utilitzarem el 80% de les dades per entrenar i el 20% per provar

# Conjunt d'entrenament (80% de les dades)
df_train = df_co2.iloc[:punt_divisio]

# Conjunt de prova (20% de les dades)
df_test = df_co2.iloc[punt_divisio:]

# Escalar les dades d'entrenament
scaler = StandardScaler()
df_train_scaled = scaler.fit_transform(df_train[['co2']])

# Escalar les dades de prova utilitzant el mateix escalar (no escalem amb el conjunt de prova)
df_test_scaled = scaler.transform(df_test[['co2']])

# Crear el model d'Isolation Forest
from sklearn.ensemble import IsolationForest
model = IsolationForest(contamination=0.1, random_state=42)

# Entrenar el model amb les dades escalades d'entrenament
model.fit(df_train_scaled)
```

Fig. 6: Codi d'entrenament del model Isolation Forest, aplicant escalat de les dades i paràmetres de contaminació.

7.5 Predicció de les Anomalies

Un cop entrenat el model, es va utilitzar per predir les anomalies en el conjunt de dades de prova (últim 20 percent de les dades). El model etiqueta les observacions com "Normal"(1) o "Anomalia"(-1), i aquests valors es van afegir al DataFrame de les dades de prova. A més, he filtrat les anomalies extremes, seleccionant només aquelles en què els nivells de CO₂ superaven els 1000 ppm, considerant-les com a potencialment perilloses.

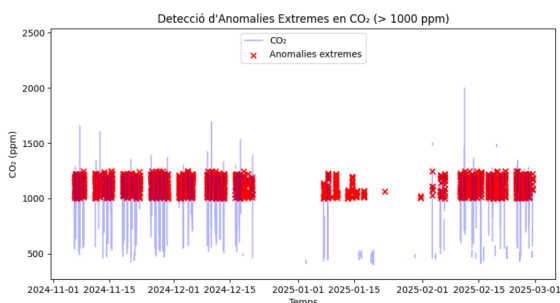


Fig. 7: Resultat del model entrenat

Com podem observar, el model funciona correctament, ja que filtra les anomalies per sobre de 1000 ppm.

7.6 Cas amb Temperatura

En aquest apartat, s'ha aplicat el mateix procediment que en l'anàlisi de CO₂, però utilitzant les dades de temperatura com a variable principal.

El model ha estat entrenat per detectar anomalies en registres de temperatura, considerant com a situació potencialment anòmla aquells valors superiors als 25°C.

A continuació, es mostren les fases de l'entrenament del model i els resultats obtinguts on es pot observar, el model és capaç de detectar valors anòmls quan la temperatura supera els 25°C, tal i com podem observar en el resultat amb les dades d'entrenament:

```
df_temperature = df_pivot[['temps', 'ubicacio', 'temperature']]

from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
df_temperature_scaled = scaler.fit_transform(df_temperature[['temperature']])

# Ordenar les dades per temps
df_temperature = df_temperature.sort_values(by='temps')

# Determinar el punt de divisió
punt_divisio = int(len(df_temperature) * 0.8) # Utilitzarem el 80% de les dades per entrenar i el 20% per provar

# Conjunt d'entrenament (80% de les dades)
df_train = df_temperature.iloc[:punt_divisio]

# Conjunt de prova (20% de les dades)
df_test = df_temperature.iloc[punt_divisio:]

# Escalar les dades d'entrenament
scaler = StandardScaler()
df_train_scaled = scaler.fit_transform(df_train[['temperature']])

# Escalar les dades de prova utilitzant el mateix escalar (no escalem amb el conjunt de prova)
df_test_scaled = scaler.transform(df_test[['temperature']])

# Crear el model d'Isolation Forest
from sklearn.ensemble import IsolationForest
model = IsolationForest(contamination=0.3, random_state=42)

# Entrenar el model amb les dades escalades d'entrenament
model.fit(df_train_scaled)
```

Fig. 8: Entrenament inicial del model amb dades de temperatura

```
df_temperature = df_pivot[['temps', 'ubicacio', 'temperature']]

from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
df_temperature_scaled = scaler.fit_transform(df_temperature[['temperature']])

# Ordenar les dades per temps
df_temperature = df_temperature.sort_values(by='temps')

# Determinar el punt de divisió
punt_divisio = int(len(df_temperature) * 0.8) # Utilitzarem el 80% de les dades per entrenar i el 20% per provar

# Conjunt d'entrenament (80% de les dades)
df_train = df_temperature.iloc[:punt_divisio]

# Conjunt de prova (20% de les dades)
df_test = df_temperature.iloc[punt_divisio:]

# Escalar les dades d'entrenament
scaler = StandardScaler()
df_train_scaled = scaler.fit_transform(df_train[['temperature']])

# Escalar les dades de prova utilitzant el mateix escalar (no escalem amb el conjunt de prova)
df_test_scaled = scaler.transform(df_test[['temperature']])

# Crear el model d'Isolation Forest
from sklearn.ensemble import IsolationForest
model = IsolationForest(contamination=0.3, random_state=42)

# Entrenar el model amb les dades escalades d'entrenament
model.fit(df_train_scaled)
```

Fig. 9: Resultat intermediari durant l'entrenament

```
# Predir les anomalies
df_test.loc[:, 'anomalies'] = model.predict(df_test_scaled)
df_test.loc[:, 'anomalies'] = df_test['anomalies'].map({1: 'Normal', -1: 'Anomalia'})

# Filtrar només les anomalies on el CO2 és superior a 1000 ppm
anomalies_extremes = df_test[(df_test['anomalies'] == 'Anomalia') & (df_test['temperature'] > 25)]

# Visualitzar només les anomalies extremes
plt.figure(figsize=(18, 5))
plt.plot(df_test['temps'], df_test['temperature'], label='temperature', color='blue', alpha=0.3)
plt.scatter(anomalies_extremes['temps'], anomalies_extremes['temperature'], color='red', label='Anomalies extremes', marker='x')
plt.title('Detecció d'Anomalies Extremes en temperatura (> 25°C)')
plt.xlabel('temps')
plt.ylabel('temperature (°C)')
plt.legend()
plt.show()
```

Fig. 10: Evolució de l'entrenament del model

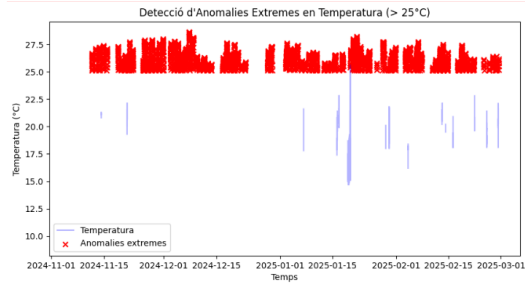


Fig. 11: Resultat del model entrenat amb dades de temperatura, identificant com a anomalies els valors superiors als 25°C.

7.7 Cas amb Humitat

Seguint la mateixa metodologia, s’ha aplicat el model sobre les dades de humitat relativa. En aquest cas, s’han considerat com a anomalies extremes aquells valors de humitat inferiors al 40% o superiors al 60%

Es mostra a continuació la visualització del resultat final del model, en què s’observen els valors classificats com a anòmals segons aquests llindars:

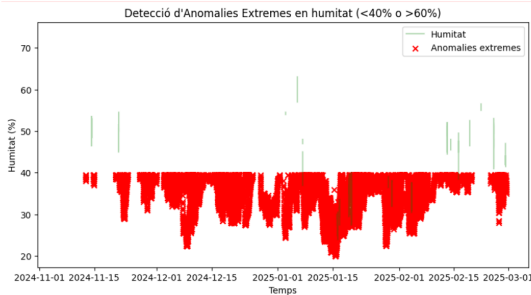


Fig. 12: Model entrenat amb detecció d’anomalies de humitat

8 RESULTATS I ANÀLISI DE DADES

Per tal de validar el model desenvolupat, s’han analitzat els resultats obtinguts en funció de les tres variables principals: CO₂, temperatura i humitat relativa. A continuació, es presenta una taula resum amb els valors estadístics més rellevants de cada variable i el nombre d’anomalies detectades pel model *Isolation Forest*.

TAULA 1: NOMBRE D’ANOMALIES EXTREMES DETECTADES PER VARIABLE AMBIENTAL

Variable	Nombre d’anomalies
CO ₂ (> 1000 ppm)	2506
Temperatura (> 25°C)	3389
Humitat (< 40% o > 60%)	29152

Els valors mostren que hi ha hagut diverses anomalies, especialment en la humitat relativa i el CO₂, cosa que reforça la necessitat d’un sistema de monitoratge intel·ligent per anticipar situacions de risc ambiental en espais tancats.

9 COMPARATIVA

TAULA 2: COMPARATIVA ENTRE SISTEMES

Aspecte	Tradicional	Isolation Forest
Detecció	Manual	Automàtica
Adaptació	Estàtica	Dinàmica
Precisió	Baixa	Alta
Escalabilitat	Limitada	Bona
Manteniment	Alt	Baix

Aquesta comparativa posa de manifest les limitacions dels sistemes manuals en entorns canviants, com els espais universitaris, on la qualitat ambiental pot variar ràpidament i sense signes evidents. En aquests casos, les solucions automatitzades permeten una detecció més eficient i proactiva, adaptant-se millor a les necessitats actuals de confort i salut. Així, l’enfocament proposat en aquest treball respon a la necessitat creixent d’eines intel·ligents per a la gestió ambiental, especialment en edificis d’ús intensiu com els educatius.

10 POTENCIAL D’APLICACIÓ DEL MODEL

El model desenvolupat amb la tècnica d’*Isolation Forest* presenta un alt potencial d’aplicació en entorns reals on es requereix monitoratge ambiental i resposta automatitzada a situacions anòmales. A diferència dels sistemes tradicionals basats en llindars predefinitos, aquest model permet identificar comportaments anòmals de forma dinàmica i intel·ligent, cosa que obre la porta a una gestió ambiental més eficient i proactiva.

10.1 Automatització de respostes mitjançant actuadors

Una de les aplicacions més prometedores del model és la seva integració amb actuadors intel·ligents. A través d’un sistema automatitzat, les anomalies detectades poden activar dispositius com ara:

- Sistemes de ventilació o finestres automàtiques, per reduir concentracions altes de CO₂.
- Sistemes de climatització, per mantenir la temperatura dins de rangs òptims.
- Humidificadors o deshumidificadors, segons els nivells de humitat relativa detectats.
- Sistemes d’il·luminació adaptativa en funció de la correlació entre presència i qualitat de l’aire.

Aquestes respostes automàtiques poden ajudar a mantenir condicions ambientals òptimes sense necessitat d’intervenció humana directa, afavorint l’eficiència energètica i el benestar dels ocupants. El que més destaca d’aquest model és que actuarà abans d’arribar a certes llindars, evitant arribar a situacions perjudicials dins de les aules.

10.2 Notificacions i manteniment predictiu

El model pot també integrar-se amb sistemes com alertes sonores o de notificació (com ara alertes per correu electrònic, missatgeria instantània o dashboards) per avisar el personal tècnic quan es detectin anomalies repetides o crítiques. Aquesta capacitat pot afavorir estratègies de manteniment predictiu i evitar situacions de risc.

10.3 Escalabilitat i adaptabilitat a altres entorns

Aquest model és escalable i adaptable a diferents tipus d'edificis, com ara escoles, oficines, laboratoris entre altres. A més, pot estendre's fàcilment a altres variables ambientals (com partícules de PM, soroll o lluminositat), simplement modificant els paràmetres d'entrada i reentrenant el model.

10.4 Millora respecte als sistemes tradicionals

La integració d'un model de detecció d'anomalies basat en *machine learning* suposa un avenç respecte als sistemes de monitoratge tradicionals. Permet una detecció precoç, una resposta contextualitzada i una gestió més intel·ligent dels recursos ambientals.

11 PROPOSTES DE MILLORA: INTEGRACIÓ DE MÉS SENSORS A LA UNIVERSITAT

11.1 Justificació de la instal·lació de més sensors intel·ligents al campus

La implementació d'una xarxa més densa de sensors ambientals a la universitat, monitoritzats mitjançant models de *machine learning* com *Isolation Forest*, suposa un salt qualitatiu respecte als sistemes tradicionals de control ambiental.

- **Millor cobertura espacial:** Amb més sensors distribuïts per les diferents facultats, laboratoris, biblioteques i espais comuns, és possible obtenir una visió més detallada i localitzada de les condicions ambientals en temps real. Això permet detectar problemes puntuals que podrien passar desapercebuts en un sistema amb pocs punts de monitoratge.
- **Resposta preventiva i automatitzada:** El model d'*Isolation Forest* permet identificar anomalies sense necessitat d'intervenció humana directa. Si s'integra amb sistemes d'actuadors (ventilació automàtica, finestres motoritzades, sistemes d'alerta, etc.), el sistema podria actuar de forma immediata davant situacions de risc, com concentracions elevades de CO₂ o temperatures extremes.
- **Millora del benestar i rendiment acadèmic:** Diversos estudis han demostrat que la qualitat de l'aire interior influeix directament en la concentració i el rendiment cognitiu. Una xarxa intel·ligent de sensors permetria garantir espais òptims per a l'estudi i la docència.
- **Escalabilitat i manteniment eficient:** Els models de *machine learning* són altament escalables. A mesura que es despleguin més sensors, el sistema pot continuar funcionant de forma eficient, millorant

continuant la seva capacitat de detecció a partir de dades reals.

- **Impacte ambiental i energètic:** Un monitoratge eficient pot ajudar a optimitzar l'ús d'energia (per exemple, millorant la ventilació segons la demanda real), reduint l'impacte ambiental a la UAB i també a altres llocs fora del campus.

11.2 Costos i estratègia d'escalabilitat

Per implementar una xarxa completa de monitoratge ambiental basada en sensors LoRaWAN, cal tenir en compte diversos costos associats. A continuació, es presenta una estimació orientativa:

- **Sensor ambiental (Milesight AM107):** aproximadament 130–150 € per unitat. Cada sensor mesura CO₂, temperatura, humitat, presència, entre altres.
- **Passarel·la LoRaWAN (gateway):** al voltant de 200–300 € per unitat, depenent de la cobertura i característiques. Les passarel·les són les encarregades de rebre les dades dels sensors i enviar-les a Internet.
- **Infraestructura de xarxa i servidors (cost inicial):** si s'utilitzen serveis al núvol gratuïts o plataformes obertes (com The Things Network), el cost pot ser mínim, però cal considerar una petita inversió en configuració i manteniment inicial (entre 100–200 €).

Tenint en compte que un sensor pot cobrir una aula mitjana, l'escenari ideal seria disposar d'un **sensor per aula i una passarel·la per cada edifici o grup d'aules**, per garantir una recepció adequada del senyal.

11.3 Estratègia d'implementació i costos

Per implementar una xarxa completa de monitoratge ambiental amb sensors LoRaWAN, cal considerar diversos costos associats, especialment en sensors, manteniment i infraestructura de comunicació. Afortunadament, l'Escola d'Enginyeria de la UAB ja disposa d'una passarel·la (gateway), fet que permet començar amb una infraestructura bàsica operativa.

El sensor utilitzat com a referència en aquest projecte, el Milesight AM107, té un cost aproximat de 120€. Aquest dispositiu mesura paràmetres com CO₂, temperatura, humitat, pressió atmosfèrica i altres indicadors de confort. A aquests costos cal afegir-ne de recurrents, com la reposició de bateries o manteniment bàsic (es pot estimar en uns 15–20€/any per sensor).

Tot i que una sola passarel·la pot cobrir diverses aules, la qualitat de la cobertura pot dependre de la distribució arquitectònica. Per garantir una cobertura òptima i escalable, seria ideal instal·lar com a mínim un sensor per aula i, si calgués, ampliar la infraestructura de passarel·les per zones amb cobertura deficient. Una passarel·la addicional té un cost aproximat de 150–200€, i pot cobrir centenars de metres en interiors.

Proposta d'escalabilitat:

- Començar desplegant sensors en espais crítics de l'Escola d'Enginyeria (aules grans, laboratoris, sales d'estudi).

- Avaluar la cobertura real de la passarel·la existent mitjançant proves pilot i, si cal, afegir-ne una segona per garantir estabilitat de connexió.
- Un cop validat el funcionament, estendre la xarxa a altres facultats o edificis del campus, de forma progressiva i prioritzant espais amb alta ocupació o ventilació deficitària.

12 DIFICULTATS

Durant el desenvolupament d'aquest sistema, s'han presentat diverses dificultats que han requerit adaptacions i solucions creatives:

- **Plantejament el tema del treball:** Inicialment, em va costar definir l'enfocament del treball, ja que, tot i tenir accés als sensors LoRaWAN de la UAB, no disposava de les dades que estaven recollint en temps real ni tenia clar com aprofitar el seu potencial. Aquesta incertesa va dificultar els primers passos, ja que no sabia ben bé quin ús pràctic podia donar als sensors dins d'un projecte acadèmic. Amb el temps, i després d'explorar diverses opcions, vaig començar a interessar-me pel concepte de detecció d'anomalies en dades ambientals, i va ser aleshores quan va sorgir la idea de desenvolupar un sistema de monitoratge intel·ligent basat en machine learning, concretament utilitzant el model *Isolation Forest*.
- **Obtenció de les dades:** Tot i haver definit finalment el propòsit del treball, un altre obstacle important va ser l'obtenció de les dades ambientals. Malgrat la presència de sensors al campus, no es va poder accedir directament a les dades que generaven. Això em va obligar a contactar amb el personal responsable de la seva gestió i esperar durant un període llarg fins que em va enviar un CSV amb totes les dades recollides durant 2 anys, ideals per entrenar el model.
- **Tractament de dades inconsistents:** El conjunt de dades inicial contenia valors nuls, formats temporals no homogenis i columnes innecessàries. Ha estat necessari aplicar diversos processos de neteja i transformació per tal de poder analitzar correctament les dades.
- **Configuració del model:** Ajustar els paràmetres del model *Isolation Forest*, ha estat clau per aconseguir una detecció realista d'anomalies. Un valor massa alt provocava massa falsos positius, i un valor massa baix, que no es detectessin anomalies reals.

Tot i aquestes dificultats, el projecte ha pogut assolir els meus objectius principals

13 CONCLUSIONS

Després de tot el que he treballat i après en aquest projecte, em queda clar que la combinació de tecnologies com LoRaWAN amb models d'intel·ligència artificial pot oferir molt més del que sovint imaginem. No es tracta només de recollir dades o fer anàlisis per curiositat acadèmica, sinó

d'aprofitar aquest coneixement per millorar, de manera real i directa, els espais on passem bona part del nostre temps.

M'he adonat que moltes vegades convivim amb problemes ambientals sense ser-ne conscients, i que podríem evitar-los si tinguéssim les eines per detectar-los a temps. Aquest projecte m'ha ajudat a veure que aquestes eines ja existeixen, són accessibles, i poden funcionar perfectament en entorns com la universitat.

Per això, crec sincerament que la nostra universitat té l'oportunitat d'anar un pas més enllà i apostar per implementar sistemes com aquest de manera més àmplia. Seria un pas cap a la innovació, i també un compromís directe amb el benestar de tota la comunitat universitària. De cara al futur, tenir sistemes d'aquest tipus combinant sensors intel·ligents i models capaços d'interpretar les dades i actuar de manera preventiva, podríem fer que els espais, sobretot dins les aules, siguin espais més sans per tothom.

Aquest projecte ha estat només una petita mostra del que és possible, però estic convençut que amb voluntat i una mirada oberta al futur, es podria convertir en una realitat molt més gran.

AGRAÏMENTS

Vull agrair especialment al meu tutor, Marc Castelló, per haver estat sempre pendent del desenvolupament d'aquest treball. Les seves orientacions constants i la dedicació durant les reunions setmanals han estat clau perquè pogués avançar amb claredat i confiança en cada fase del projecte.

També voldria donar les gràcies a Marc Codina, per haver-me facilitat les dades ambientals que han fet possible la part pràctica del treball, especialment l'entrenament del model. Sense aquest suport tècnic, hauria estat molt difícil portar a terme la part d'anàlisi que volia fer.

Finalment, agraeixo a Juan Vicente i Alfons Miàs per haver-me introduït i despertat l'interès per aquest tipus de tecnologies. Gràcies a ells vaig descobrir un camp que m'ha acabat apassionant i que, sens dubte, m'ha marcat tant en aquest projecte com en la meua trajectòria com a estudiant.

REFERÈNCIES

- [1] Satish, U., Mendell, M. J., Shekhar, K., Hotchi, T., Sullivan, D., Streufert, S., & Fisk, W. J. (2012). Is CO2 an indoor pollutant? Direct effects of low-to-moderate CO2 concentrations on human decision-making performance. *Environmental Health Perspectives**, 120(12), 1671–1677.
- [2] Wargocki, P., & Wyon, D. P. (2007). The effects of moderately raised classroom temperatures and classroom ventilation rate on the performance of schoolwork by children (RP-1257). *HVAC&R Research**, 13(2), 193–220.

- [3] Arundel, A. V., Sterling, E. M., Biggin, J. H., & Sterling, T. D. (1986). Indirect health effects of relative humidity in indoor environments. **Environmental Health Perspectives**, 65, 351–361.
- [4] Seppänen, O. A., Fisk, W. J., & Mendell, M. J. (1999). Association of ventilation rates and CO₂ concentrations with health and other responses in commercial and institutional buildings. **Indoor Air**, 9(4), 226–252.
- [5] Wyon, D. P. (2004). The effects of indoor air quality on performance and productivity. **Indoor Air**, 14(s7), 92–101.
- [6] Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008, December). Isolation forest. In **2008 Eighth IEEE International Conference on Data Mining** (pp. 413–422). IEEE.
- [7] Binetti, M. S., Uricchio, V. F., & Massarelli, C. (2025). Isolation Forest for Environmental Monitoring: A Data-Driven Approach to Land Management. **Environments**, 12(4), 116.
- [8] Thu, M. Y., Htun, W., Aung, Y. L., Shwe, P. E. E., & Tun, N. M. (2018, November). Smart air quality monitoring system with LoRaWAN. In **2018 IEEE International Conference on Internet of Things and Intelligence System (IOTAIS)** (pp. 10–15). IEEE.
- [9] Ali, S., Glass, T., Parr, B., Potgieter, J., & Alam, F. (2020). Low-cost sensor with IoT LoRaWAN connectivity and machine learning-based calibration for air pollution monitoring. **IEEE Transactions on Instrumentation and Measurement**, 70, 1–11.
- [10] González, E., Casanova-Chafer, J., Romero, A., Vilanova, X., Mitrovics, J., & Llobet, E. (2020). LoRa sensor network development for air quality monitoring or detecting gas leakage events. **Sensors**, 20(21), 6225.
- [11] Wang, C., Vo, H. T., & Ni, P. (2015, December). An IoT application for fault diagnosis and prediction. In *2015 IEEE International Conference on Data Science and Data Intensive Systems* (pp. 726–731). IEEE.
- [12] Patrão, R. L., de Caldas Filho, F. L., e Martins, L. M., Silva, G. D. N., Monteiro, M. S., Andrade, M. B., ... & de Sousa Junior, R. T. (2020, June). Environmental building monitoring and control based on machine learning and fog computing on an IoT architecture. In *2020 15th Iberian Conference on Information Systems and Technologies (CISTI)* (pp. 1–6). IEEE.
- [13] Torres, W., Sánchez, J., Bustos, S., & Coronel, M. (2024). Arquitectura de IoT para el monitoreo de emisiones de gases contaminantes de vehículos y su validación a través de Machine Learning.