**Disclaimer**

This content has been generated by neural machine translation or generative AI. The translations are automated and have not undergone human review or validation. If you plan to distribute automatically translated content, we suggest you review it in advance for translation accuracy and formatting.

# Previsione basata su scenari e mitigazione dei rischi ambientali

## Risultati apprendimento

Dopo aver completato questo argomento, gli studenti saranno in grado di:

- Spiegare il concetto di previsione degli scenari nell'analisi dei dati ambientali e climatici.
- Descrivere il processo di aggregazione delle serie temporali di dati ambientali in indicatori climatici annuali e spiegare il suo ruolo nell'analisi dei rischi a lungo termine.
- Applicare il metodo di apprendimento automatico non supervisionato (K-means) per raggruppare gli anni climatici in base alle loro caratteristiche generali.
- Interpretare i cluster risultanti come diversi scenari climatici
- Discutere come i risultati dell'analisi degli scenari possono supportare il processo decisionale

## Introduzione

I moderni sistemi di monitoraggio ambientale dipendono sempre più dalla raccolta continua di dati provenienti da varie fonti, tra cui reti di sensori, stazioni meteorologiche, osservazioni satellitari e archivi ecologici aperti. Questi sistemi generano regolarmente grandi volumi di dati e costituiscono una base per il monitoraggio delle condizioni ambientali su lunghi periodi di tempo.

A differenza dei metodi analitici tradizionali, che spesso si basano su set di dati statici o limitati, il monitoraggio ambientale è un processo dinamico in cui i dati vengono raccolti e aggiornati continuamente. In questo contesto, il machine learning implica non solo lo sviluppo di singoli modelli, ma anche il mantenimento di un approccio analitico continuo per gestire i dati strutturati nel tempo ed eseguire analisi a lungo termine.

Le variabili ambientali più importanti, come i livelli di inquinanti atmosferici, la temperatura, le precipitazioni o il livello del mare, sono naturalmente organizzate in serie temporali. Questi dati mostrano la stagionalità, le fluttuazioni a breve termine e le tendenze a lungo termine, che richiedono metodi che analizzano come le condizioni cambiano nel tempo, anziché semplicemente osservare gli stati attuali.

In questo contesto, il machine learning offre metodi per monitorare le tendenze e rilevare anomalie nei dati ambientali. Per essere praticamente utili nei sistemi operativi, questi metodi devono essere integrati in un più ampio framework di gestione dei dati ecologici piuttosto che utilizzati come strumenti analitici isolati, a supporto di decisioni ambientali informate e responsabili come la diagnosi precoce dei problemi di qualità dell'aria, la valutazione del rischio di eventi meteorologici estremi o il monitoraggio dei cambiamenti del livello del mare a lungo termine.

## Dal monitoraggio ambientale continuo agli scenari climatici

Nel monitoraggio ambientale a lungo termine, basarsi esclusivamente su misurazioni individuali o su brevi periodi di tempo è spesso insufficiente. Per rilevare cambiamenti significativi dei rischi climatici, i dati sono generalmente riassunti su periodi più lunghi, come l'anno. Questo metodo consente ai singoli anni di fungere da istantanee complete delle condizioni ambientali, anziché soltanto come misurazioni isolate.

Oltre ai valori medi, eventi estremi come ondate di calore o incantesimi secchi sono cruciali nell'analisi del rischio ambientale. Il loro significato deriva non solo dalla loro intensità, ma anche dalla frequenza con cui si verificano e dal tempo che dure. Pertanto, l'analisi considera non solo le temperature medie e le precipitazioni, ma anche la frequenza, la durata e le combinazioni di condizioni dannose.

Quando questi indicatori vengono considerati insieme, ogni anno può essere caratterizzato da una serie di tratti che definiscono il suo profilo climatico. In questo contesto, le tecniche di machine learning possono essere utilizzate per identificare somiglianze tra gli anni e raggrupparle in base ai tratti complessivi.

## Algoritmo di clustering K-Means nel contesto dei dati ambientali

K-means [1], [2], [3] è un algoritmo di clustering semplice e ampiamente utilizzato per identificare somiglianze in set di dati di grandi dimensioni. La sua idea principale è quella di suddividere i dati in un numero fisso di gruppi (cluster), in cui ogni gruppo contiene record più simili tra loro in base alle caratteristiche osservate. In questo metodo, ogni anno è rappresentato da una serie di caratteristiche numeriche e l'algoritmo esamina contemporaneamente tutte queste caratteristiche, cercando la somiglianza complessiva tra i record anziché la corrispondenza in una singola variabile.

In pratica, l'algoritmo parte dalle stime iniziali dei centri del gruppo. Quindi li regola ripetutamente per ridurre la distanza tra ciascun record e il suo centroide di gruppo. Questo

processo prosegue finché la distribuzione dei documenti tra i gruppi si stabilizza. Il risultato sono gruppi che rappresentano pattern comuni nei dati.

Nelle analisi ambientali, K-means raggruppa diversi periodi di tempo, come gli anni singoli, in base al loro profilo climatico complessivo anziché a una sola variabile. Questo metodo aiuta a identificare condizioni climatiche stabili, transitorie ed estreme senza fare affidamento su soglie o regole predefinite.

Immaginiamo che ogni anno tracciamo alcuni semplici indicatori come la temperatura media, il numero di giorni molto caldi e i giorni senza pioggia. Considerando singolarmente questi indicatori, ogni anno sembra unico, il che rende difficile descriverli con chiarezza.

Tuttavia, quando tutti gli indicatori vengono considerati insieme, alcuni anni presentano modelli molto simili, come mostrato nella Figura 5.1. Ad esempio, in alcuni anni si possono registrare temperature moderate, ondate di calore rare e precipitazioni sufficienti. La seconda metà dell'anno tende a registrare temperature più elevate, ondate di calore più frequenti e periodi più secchi.

In questo caso, l'algoritmo K-means non specifica in anticipo il numero di gruppi; invece, individua automaticamente gruppi di anni più simili tra tutti gli indicatori osservati. Ogni gruppo risultante può essere considerato un modello climatico tipico.

Un gruppo può rappresentare anni più stabili e moderati; il secondo può indicare uno stress termico significativo; il terzo potrebbe essere collegato a calore estremo e a siccità. Quando si analizza un nuovo set di dati, ogni anno può essere confrontato con questi modelli e classificato nel gruppo che assomiglia più strettamente in base alle sue caratteristiche.

Il caso di studio descritto di seguito si basa su questo concetto, utilizzando serie temporali meteorologiche a lungo termine per classificare gli anni climatici effettivi e analizzarli nel contesto dei rischi climatici.

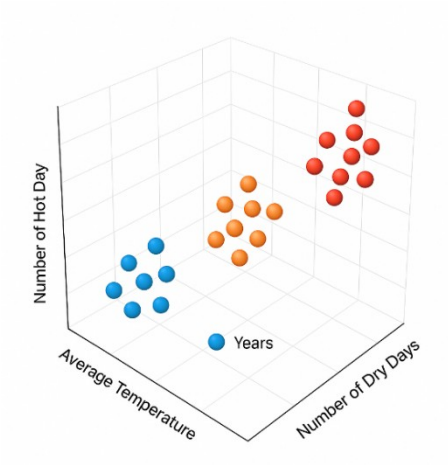


Figura 5.1 Esempio illustrativo di clustering dei profili climatici annuali

## Caso di studio: Analisi del rischio climatico basata sulle serie meteorologiche

### Origine dati Open-Meteo

Nel caso di studio, la piattaforma Open-Meteo [4] è utilizzata come riferimento per i dati meteo. Open-Meteo fornisce l'accesso gratuito ai dati meteorologici storici per qualsiasi posizione in tutto il mondo tramite un'API pubblica, consentendo agli utenti di scegliere un intervallo di tempo e una serie di variabili meteorologiche. I dati storici risalgono al 1940, con un leggero ritardo (fino a 1 giorno) nell'accesso ai record più recenti.

Mentre l'accesso ai dati di base è gratuito, Open-Meteo limita il modello di applicazione aperto limitando le chiamate API e le funzionalità avanzate. Per gli utenti professionali e aziendali sono disponibili piani API estesi che offrono limiti più elevati, un maggior numero di set di dati e un'infrastruttura più affidabile.

I dati di Open-Meteo provengono da risorse pubbliche, tra cui le agenzie meteorologiche nazionali e la rianalisi climatica. È fornito con la licenza Creative Commons Attribuzione 4.0 (CC BY 4.0), che impone agli utenti di attribuire la fonte. Questa licenza consente il suo utilizzo nel lavoro educativo e di ricerca, promuovendo al contempo la trasparenza e la riproducibilità dell'analisi.

Il download dovrebbe essere eseguito tramite un'interfaccia di rete semplice, specificando la posizione geografica, l'intervallo di tempo (Figura 5.2), la risoluzione temporale (ore o giornaliera) e le variabili meteorologiche selezionate. Questo metodo garantisce una raccolta accurata e coerente dei dati in diverse posizioni e in diversi scenari di analisi.

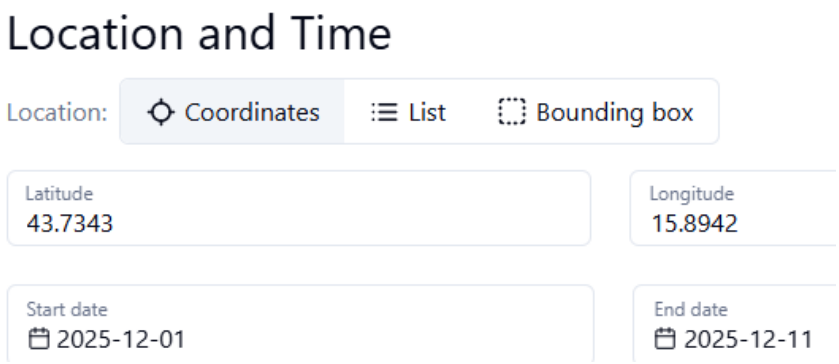


Figura 5.2 Selezione della posizione dei dati e della serie temporale sulla piattaforma Open-Meteo

## Esempio di Clustering Climate Years

Per l'analisi del rischio climatico, è importante selezionare solo le variabili direttamente correlate ai cambiamenti a lungo termine e allo stress ambientale, il cui significato è chiaramente comprensibile. In questo contesto, il focus sarà sulle variabili che descrivono le condizioni termiche, l'equilibrio idrologico e gli eventi meteorologici estremi.

Gli indicatori di base dello stress termico comprendono la temperatura media giornaliera dell'aria a 2 metri (temperatura media (2 m) – temperature_2m_mean) e la temperatura massima giornaliera dell'aria (temperatura massima (2 m) – temperature_2m_max). Queste misurazioni aiuteranno a monitorare le tendenze di riscaldamento a lungo termine e a identificare i periodi di calore estremi. Inoltre, la temperatura media apparente (2 m) (apparent_temperature_mean), che rappresenta gli effetti dell'umidità e del vento, sarà utilizzata per fornire una valutazione più accurata del carico termico sull'uomo e sull'ambiente.

Per valutare lo stress idrologico e quello secco, saranno scelte variabili per descrivere la disponibilità e la perdita di acqua nell'ambiente. La precipitazione giornaliera totale (Precipitation Sum – precipitation_sum) sarà utilizzata per stimare l'ingresso di acqua nel sistema, mentre l'evapotraspirazione di riferimento (Reference Evapotranspiration – et0_fao_evapotranspiration) rappresenterà la potenziale perdita di acqua dovuta all'evaporazione e al trasporto. L'umidità media del suolo nello strato 0-100 cm (umidità media del suolo (0-100 cm) - soil_moisture_0_100cm_mean) servirà come indicatore della disponibilità effettiva di acqua nel terreno e dello stress secco prolungato.

Come variabili supplementari che forniscono un contesto meteorologico più ampio e aiutano a interpretare i risultati, verranno utilizzate la velocità massima del vento (Velocità massima del

vento (10 m) – wind_speed_10m_max), l'umidità relativa media (Umidità relativa media (2 m) – relative_humidity_2m_mean) e la pressione media dell'aria ridotta al livello del mare (Pressione del livello del Mar Mediterraneo – pressure_msl_mean). Anche se queste variabili non sono centrali per l'analisi, miglioreranno la comprensione delle condizioni in cui i rischi climatici si sviluppano e variano.

## Periodo di osservazione e ambito temporale dell'analisi

I dati meteorologici storici dal 1° gennaio 1950 al 31 dicembre 2024 saranno utilizzati per l'analisi del rischio climatico. Il periodo scelto fornisce una serie temporale sufficientemente lunga per identificare i cambiamenti a lungo termine delle condizioni termiche e idrologiche, mantenendo una qualità dei dati relativamente alta e stabile, essenziale per un'analisi e un'interpretazione accurate.

I dati verranno scaricati quotidianamente, bilanciando la risoluzione temporale e la stabilità della serie. La risoluzione giornaliera aiuta a monitorare i modelli stagionali e i cambiamenti nella frequenza delle condizioni avverse, riducendo al contempo il rumore e le fluttuazioni a breve termine presenti nei dati a più alta frequenza.

## Flusso di lavoro analitico e uso dei metodi di apprendimento automatico

Nell'analisi degli indicatori termici e idrologici, l'attenzione non si concentrerà semplicemente sulla presenza di modelli stagionali, attesi e ben noti (ad esempio i periodi asciutti estivi), ma piuttosto su come le loro caratteristiche varino su periodi più lunghi. In particolare, l'analisi esaminerà i cambiamenti nella frequenza, nella durata e nell'intensità delle condizioni avverse, insieme a possibili variazioni nei loro tempi entro l'anno.

Nell'ambito dello stress idrologico, l'analisi si concentrerà sui periodi in cui la disponibilità di acqua del suolo diminuisce per diversi anni consecutivi, in combinazione con un aumento dell'evapotraspirazione. Valuterà anche se questi periodi si verificano prima della stagione, durano più a lungo o mostrano una maggiore intensità rispetto agli anni precedenti. Questi cambiamenti indicano un aumento del rischio di siccità, anche se il modello stagionale del fenomeno resta invariato.

Allo stesso modo, gli indicatori di calore non solo monitoreranno i massimi estivi, ma anche i cambiamenti nel numero di giorni caldi, nella durata dei periodi di calore e nella frequenza delle alte temperature al di fuori dei tipici intervalli stagionali. Ciò aiuta a differenziare i modelli stagionali naturali dai turni a lungo termine, con implicazioni dirette per i rischi climatici.

## Importazione dati in un data warehouse Oracle

I dati meteorologici scaricati vengono importati tramite l'interfaccia Data Load incorporata in Oracle Autonomous Data Warehouse [5]. L'interfaccia consente di caricare i dati da file .xlsx locali e di convertirli automaticamente in tabelle relazionali all'interno del database, come mostrato nella Figura 5.3.

Durante il caricamento, viene specificato il nome della tabella di destinazione e le colonne del file Excel vengono mappate ai relativi tipi di dati del database. I dati vengono memorizzati come attributi temporali, mentre le variabili meteorologiche numeriche vengono caricate come valori numerici adatti a ulteriori analisi, come mostrato nella Figura 5.4.

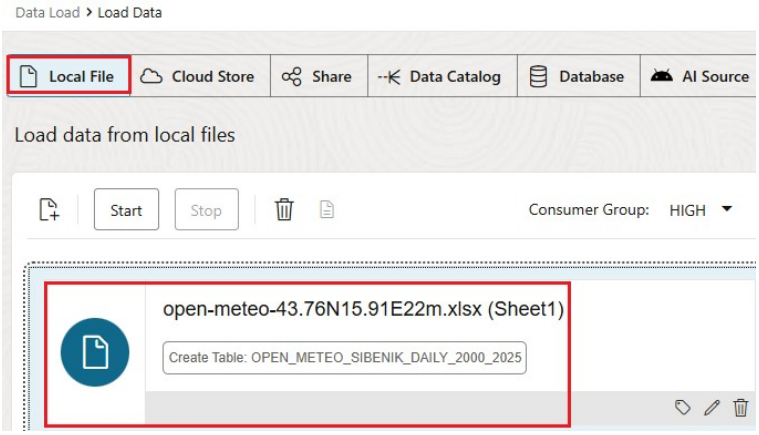


Figura 5.3 Interfaccia di caricamento dati per l'importazione di dati meteorologici in formato XLSX

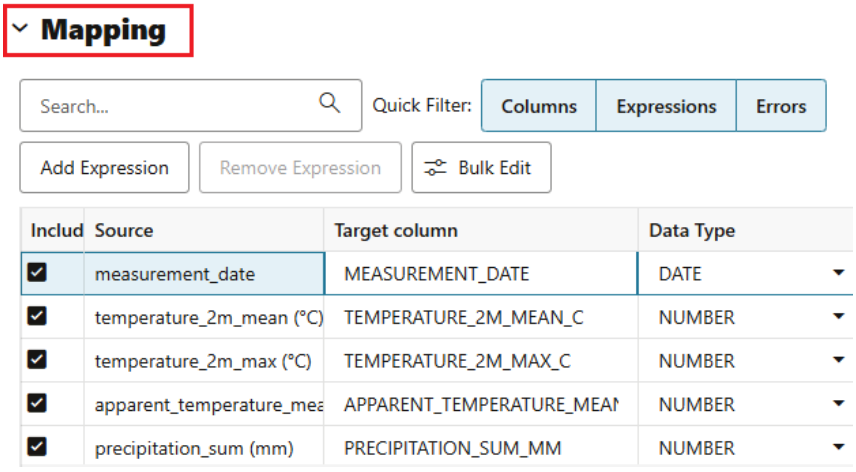


Figura 5.4 Mapping delle colonne da un file di Excel ai tipi di dati del database durante l'importazione

Questa procedura si traduce in una tabella strutturata con serie temporali giornaliere di variabili meteorologiche selezionate per la posizione osservata (Šibenik) e per il periodo specificato (dal 01/01/1940 al 31/12/2024), come illustrato nella figura 5.5.

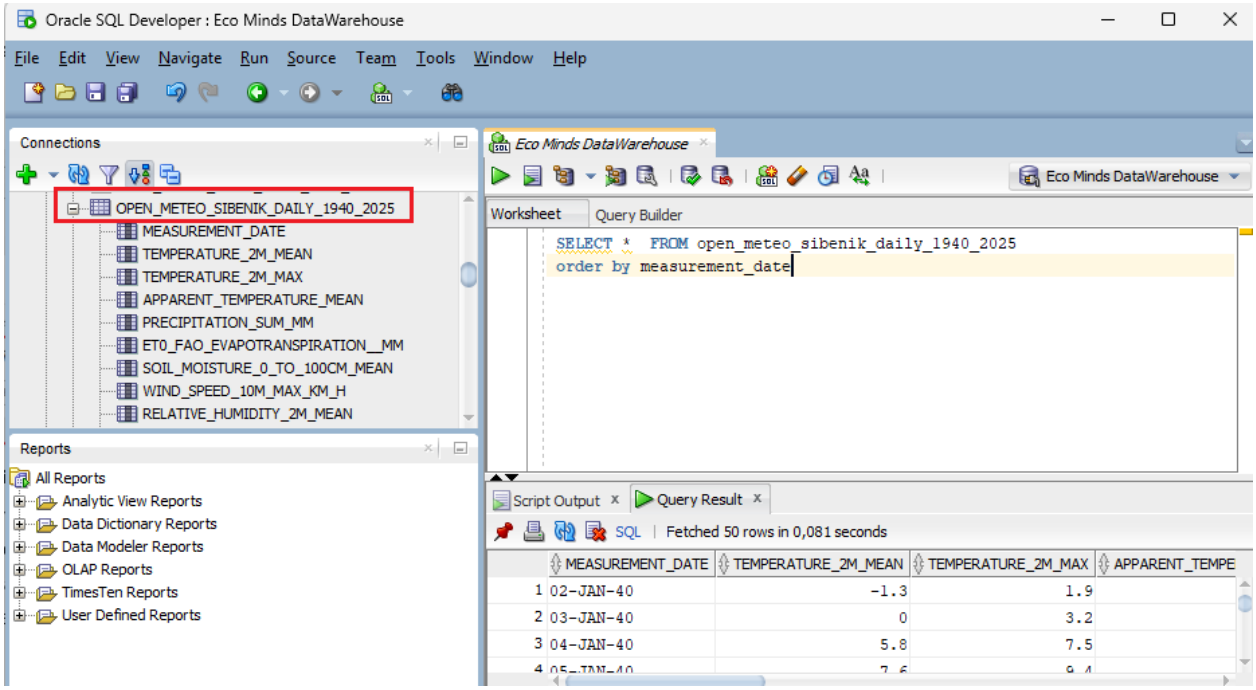


Figura 5.5 Struttura della tabella OPEN_METEO_SIBENIK_DAILY_1940_2025 con dati meteorologici giornalieri

## Aggregazione dei dati e progettazione delle funzioni

Dopo aver importato le misurazioni meteorologiche giornaliere nel database relazionale, il passo successivo è aggregare i dati e creare funzionalità di input annuali (un insieme di funzioni). Poiché l'obiettivo dell'analisi è identificare modelli a lungo termine e tipi di anni climatici, i record giornalieri vengono convertiti in riepiloghi annuali che caratterizzano le condizioni termiche, idrologiche ed estreme per ogni anno, formando un unico record multidimensionale.

L'aggregazione viene eseguita tramite una query SQL che utilizza più CTE (espressioni di tabella comune). Ogni parte calcola un set specifico di indicatori, e i risultati vengono riuniti in un'unica tabella annuale delle caratteristiche.

Nel blocco **yearly_temps**, l'immagine termica annuale viene generata utilizzando misure statistiche della temperatura dell'aria a 2 metri e della temperatura della pelle. Ogni anno conta::

- temperatura media media giornaliera e mediana (AVG, PERCENTILE_CONT(0.5)), coprendo contemporaneamente una temperatura "tipica" e una misura robusta meno sensibile agli estremi,
- valori minimi e massimi (MIN, MAX), se il massimo è prelevato da temperature_2m_max per catturare meglio il carico termico estremo,
- analoghe misurazioni della temperatura di rilevamento (apparent_temperature_mean) come indicatore di stress termico effettivo, che comprende l'impatto dell'umidità e del vento,
- Inoltre, l'umidità relativa media è calcolata come variabile contestuale che aiuta a interpretare le differenze tra il "calore secco" e i periodi con maggiore carico termico dovuto all'umidità.

Gli eventi di calore estremo vengono valutati in base ai criteri dell'onda di calore. Nella sezione onda di calore, i giorni vengono identificati per la prima volta quando temperature_2m_max supera i 30 °C e sono consecutivi tra loro. La continuità viene stabilita raggruppando le date consecutive in base alla differenza tra la data e il numero di sequenza. Solo gli episodi di durata almeno 5 giorni (HAVING COUNT(*) >= 5) sono classificati come ondate di calore. Utilizzando questi episodi, heatwave_stats calcola i totali annuali.

- il numero di ondate di calore nell'anno,
- durata media, minima e massima dell'ondata di calore.

Oltre al livello di temperatura stesso, la struttura degli eventi estremi (come la frequenza con cui si verificano e la durata) viene catturata, che è vitale per comprendere il rischio climatico.

Per caratterizzare il rischio secco come integrato (e non solo tramite le precipitazioni), nel set di dati yearly_drought è stato introdotto un indicatore composito, "drought risk day". Tale giorno è definito come il simultaneo adempimento di tre condizioni:

- precipitazioni molto basse (precipitation_sum_mm  <1),
- evapotraspirazione di riferimento elevato (et0_fao_evapotranspiration__mm >4),
- bassa umidità del suolo (soil_moisture_0_to_100cm_mean < 0,20).

Inoltre, vengono identificati giorni specifici con elevata evapotraspirazione e bassa umidità del suolo, consentendo di differenziare tra anni in cui domina la "domanda atmosferica di acqua" e quelli in cui la questione principale è un deficit idrico a lungo termine nel terreno. In questo modo, lo stress secco non viene eccessivamente semplificato in un unico segnale. Invece, è modellato in funzione di input (precipitazione), di perdite (ET0) e della disponibilità effettiva di acqua (umidità del suolo).

Il blocco **yearly_wind** aggiunge un componente a condizioni di vento estreme tramite:

- velocità massima del vento registrata in un anno,
- Numero di giorni con vento forte (soglia > 50 km/h).

Questi indicatori non sono il principale focus del rischio termico-idrologico, ma contribuiscono a comprendere gli eventi estremi rilevanti per il monitoraggio operativo e la gestione del rischio.

Il SELECT finale combina tutti i set di funzionalità annuali calcolati in un unico risultato, utilizzando LEFT JOIN per gli indicatori che non devono esistere ogni anno (ad esempio, le ondate di calore), con i valori mancanti riempiti con 0 (COALESCE). L'output è un set di dati annuale delle funzionalità di input, con ogni anno come record e molteplici fattori di stress ambientale: calore, precipitazioni, siccità e vento. Questa matrice di funzioni rappresenta lo spazio di input per l'uso successivo di metodi di apprendimento automatico non supervisionati (ad esempio, K-means).

```
WITH yearly_temps AS (
  SELECT
    EXTRACT(YEAR FROM measurement_date) AS year,

    -- Air temperature statistics
    ROUND(AVG(temperature_2m_mean), 1) AS avg_year_temp,
    ROUND(
      PERCENTILE_CONT(0.5)
      WITHIN GROUP (ORDER BY temperature_2m_mean),
      1
    ) AS median_year_temp,
    MIN(temperature_2m_mean) AS min_year_temp,
    MAX(temperature_2m_max) AS max_year_temp,

    -- Apparent temperature (thermal stress)
    ROUND(AVG(apparent_temperature_mean), 1) AS avg_apparent_temp,
    ROUND(
      PERCENTILE_CONT(0.5)
      WITHIN GROUP (ORDER BY apparent_temperature_mean),
      1
    ) AS median_apparent_temp,
    MIN(apparent_temperature_mean) AS min_apparent_temp,
    MAX(apparent_temperature_mean) AS max_apparent_temp,

    -- Relative humidity
    ROUND(AVG(relative_humidity_2m_mean), 1) AS avg_relative_humidity

  FROM open_meteo_sibenik_daily_1940_2025
  WHERE measurement_date BETWEEN
        DATE '1950-01-01' AND DATE '2024-12-31'
  GROUP BY EXTRACT(YEAR FROM measurement_date)
),
yearly_hydrology AS (
```

```
  SELECT
    EXTRACT(YEAR FROM measurement_date) AS year,

    -- Hydrological indicators
    COUNT(CASE WHEN precipitation_sum_mm < 1 THEN 1 END) AS dry_days,
    COUNT(CASE WHEN precipitation_sum_mm >= 1 THEN 1 END) AS rainy_days,
    MAX(precipitation_sum_mm) AS max_daily_precipitation_mm

  FROM open_meteo_sibenik_daily_1940_2025
  WHERE measurement_date BETWEEN
        DATE '1950-01-01' AND DATE '2024-12-31'
  GROUP BY EXTRACT(YEAR FROM measurement_date)
),
heatwaves AS (
  SELECT
    grp,
    MIN(measurement_date) AS start_date,
    COUNT(*) AS duration_days
  FROM (
    SELECT
      measurement_date,
      measurement_date
        - ROW_NUMBER() OVER (ORDER BY measurement_date) AS grp
    FROM open_meteo_sibenik_daily_1940_2025
    WHERE measurement_date BETWEEN
          DATE '1950-01-01' AND DATE '2024-12-31'
      AND temperature_2m_max > 30
  )
  GROUP BY grp
  HAVING COUNT(*) >= 5
),
heatwave_stats AS (
  SELECT
    EXTRACT(YEAR FROM start_date) AS year,

    -- Heatwave indicators
    COUNT(*) AS heatwave_count,
    ROUND(AVG(duration_days), 1) AS heatwave_avg_duration_days,
    MIN(duration_days) AS heatwave_min_duration_days,
    MAX(duration_days) AS heatwave_max_duration_days

  FROM heatwaves
  GROUP BY EXTRACT(YEAR FROM start_date)
),
yearly_drought AS (
  SELECT
    EXTRACT(YEAR FROM measurement_date) AS year,

    -- Combined drought risk (all conditions met)
    COUNT(
      CASE
        WHEN precipitation_sum_mm < 1
         AND et0_fao_evapotranspiration__mm > 4
```

```
          AND soil_moisture_0_to_100cm_mean < 0.20
        THEN 1
      END
    ) AS drought_risk_days,

    -- Individual drought-related conditions
    COUNT(
      CASE
        WHEN et0_fao_evapotranspiration__mm > 4
        THEN 1
      END
    ) AS high_et0_days,

    COUNT(
      CASE
        WHEN soil_moisture_0_to_100cm_mean < 0.20
        THEN 1
      END
    ) AS low_soil_moisture_days

  FROM open_meteo_sibenik_daily_1940_2025
  WHERE measurement_date BETWEEN
        DATE '1950-01-01' AND DATE '2024-12-31'
  GROUP BY EXTRACT(YEAR FROM measurement_date)
),
yearly_wind AS (
  SELECT
    EXTRACT(YEAR FROM measurement_date) AS year,

    -- Wind extremes
    MAX(wind_speed_10m_max_km_h) AS max_wind_speed_km_h,
    COUNT(
      CASE
        WHEN wind_speed_10m_max_km_h > 50
        THEN 1
      END
    ) AS strong_wind_days_50kmh

  FROM open_meteo_sibenik_daily_1940_2025
  WHERE measurement_date BETWEEN
        DATE '1950-01-01' AND DATE '2024-12-31'
  GROUP BY EXTRACT(YEAR FROM measurement_date)
)
SELECT
  t.year,

  -- Heatwave indicators
  COALESCE(h.heatwave_count, 0) AS heatwave_count,
  COALESCE(h.heatwave_avg_duration_days, 0) AS heatwave_avg_duration_days,
  COALESCE(h.heatwave_min_duration_days, 0) AS heatwave_min_duration_days,
  COALESCE(h.heatwave_max_duration_days, 0) AS heatwave_max_duration_days,

  -- Air temperature
```

```
  t.avg_year_temp,
  t.median_year_temp,
  t.min_year_temp,
  t.max_year_temp,

  -- Apparent temperature
  t.avg_apparent_temp,
  t.median_apparent_temp,
  t.min_apparent_temp,
  t.max_apparent_temp,

  -- Relative humidity
  t.avg_relative_humidity,

  -- Hydrology
  y.dry_days,
  y.rainy_days,
  y.max_daily_precipitation_mm,

  -- Drought stress
  COALESCE(d.drought_risk_days, 0) AS drought_risk_days,
  COALESCE(d.high_et0_days, 0) AS high_et0_days,
  COALESCE(d.low_soil_moisture_days, 0) AS low_soil_moisture_days,

  -- Wind
  w.max_wind_speed_km_h,
  w.strong_wind_days_50kmh

FROM yearly_temps t
LEFT JOIN heatwave_stats h ON t.year = h.year
JOIN yearly_hydrology y ON t.year = y.year
LEFT JOIN yearly_drought d ON t.year = d.year
LEFT JOIN yearly_wind w ON t.year = w.year
ORDER BY t.year;
```

## Analisi e visualizzazione di indicatori annuali aggregati

Sulla base dei totali annuali calcolati dalla query SQL descritta nel precedente sottocapitolo, è stata condotta un'analisi visiva dei principali indicatori termici e idrologici. Questo funge da strumento analitico supplementare per comprendere i modelli a lungo termine e monitorare i cambiamenti nelle condizioni climatiche nel tempo.

La figura 5.6 mostra la temperatura media annua dell'aria a Sebenico dal 1950 al 2024, calcolata come media annua della temperatura media giornaliera a 2 m (temperature_2m_mean).

La visualizzazione conferma una tendenza a lungo termine all'aumento delle temperature, con marcata variabilità anno per anno. In particolare, dall'inizio del 2000 le temperature medie sono state superiori alle medie pluriennali precedenti e, negli ultimi anni, si sono registrati periodi di temperature significativamente elevate.

Questa tendenza suggerisce non solo un riscaldamento graduale, ma anche un cambiamento nella distribuzione delle condizioni termiche annuali, fondamentale per l'analisi delle onde di calore e per raggruppare gli anni climatici mediante metodi di apprendimento automatico.

La figura 5.7 mostra il numero annuo di ondate di calore, definite come periodi di almeno cinque giorni consecutivi con una temperatura massima giornaliera superiore a 30 °C, insieme alla loro durata media. All'inizio della serie, le ondate di calore sono rare e di norma brevi. Tuttavia, negli ultimi anni la loro frequenza è aumentata e la durata dei singoli eventi è aumentata in modo significativo.

Questo schema mostra un cambiamento nel modo in cui lo stress termico ci influenza: anziché eventi estremi occasionali, periodi più lunghi di temperature elevate sono diventati più frequenti negli ultimi anni, influenzando in modo significativo la salute umana, gli ecosistemi e l'uso delle risorse.

La figura 5.8 mostra il conteggio annuale dei giorni secchi e piovosi, con i giorni secchi definiti come quelli con precipitazioni inferiori a 1 mm. Il grafico rivela una notevole variazione anno per anno, con una tendenza recente verso giorni più secchi e meno giorni di pioggia.

Questo cambiamento nella distribuzione dei giorni di precipitazione non implica necessariamente una diminuzione delle precipitazioni annuali totali. Invece, mostra una maggiore concentrazione di eventi di pioggia e di periodi più lunghi di secchezza, entrambi importanti per lo sviluppo dello stress secco.

La figura 5.9 illustra il numero annuale di giorni che soddisfano i criteri per un rischio secco pronunciato, caratterizzato da precipitazioni molto basse, un aumento dell'evapotraspirazione di riferimento e una riduzione dell'umidità del suolo. Durante la maggior parte del periodo osservato, tali condizioni non sono soddisfatte contemporaneamente; tuttavia, negli ultimi anni, ci sono stati giorni notevolmente più secchi.

L'emergere di periodi pluridecennali di stress arido combinato indica un cambiamento nel modo in cui i fattori atmosferici e idrologici interagiscono, con la siccità non più dovuta solo alle precipitazioni, ma anche a una maggiore perdita di acqua e a una minore umidità del suolo.

È importante riconoscere che le caratteristiche avverse non si verificano necessariamente ogni anno; le loro combinazioni e i livelli di espressione variano. La natura multidimensionale del clima annuale sottolinea la necessità di un metodo in grado di analizzare più indicatori contemporaneamente e di rilevare modelli naturali nel loro verificarsi.

A seguito di ciò, il passo successivo nell'analisi consiste nell'utilizzare metodi di apprendimento automatico non supervisionati che trattano i record annuali come profili climatici completi. Questo approccio consente di raggruppare gli anni in base alle loro somiglianze generali, contribuendo a identificare diversi tipi di anni climatici e interpretarli in termini di rischi climatici.

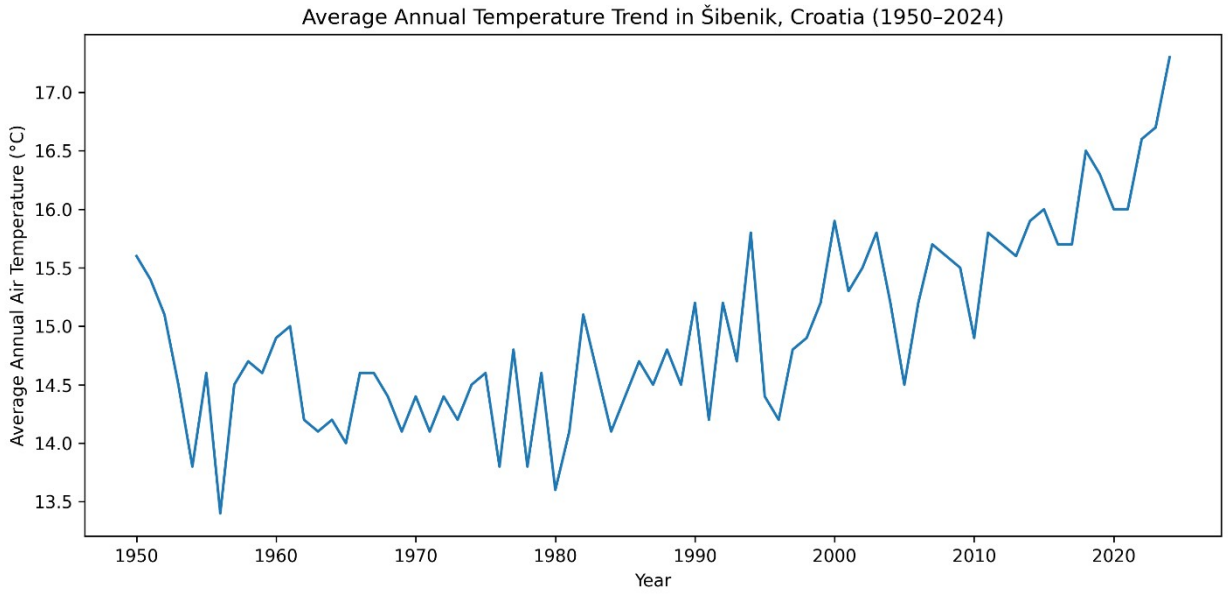


Figura 5.6 Temperatura media annuale dell'aria a Sebenico (1950-2024)

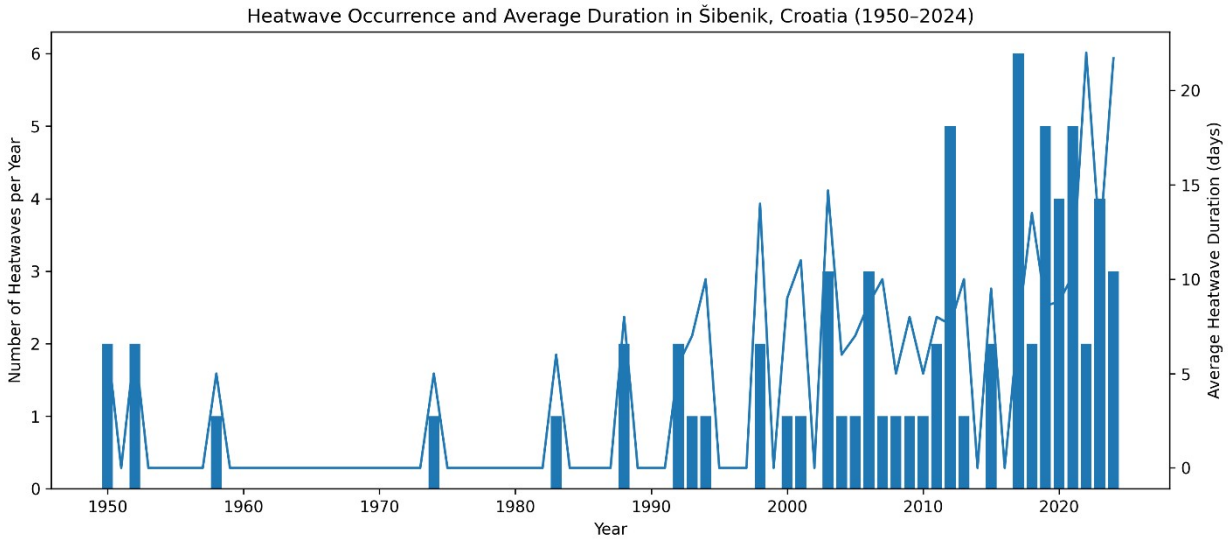


Figura 5.7 Frequenza e durata media delle onde di calore a Sebenico (1950-2024)

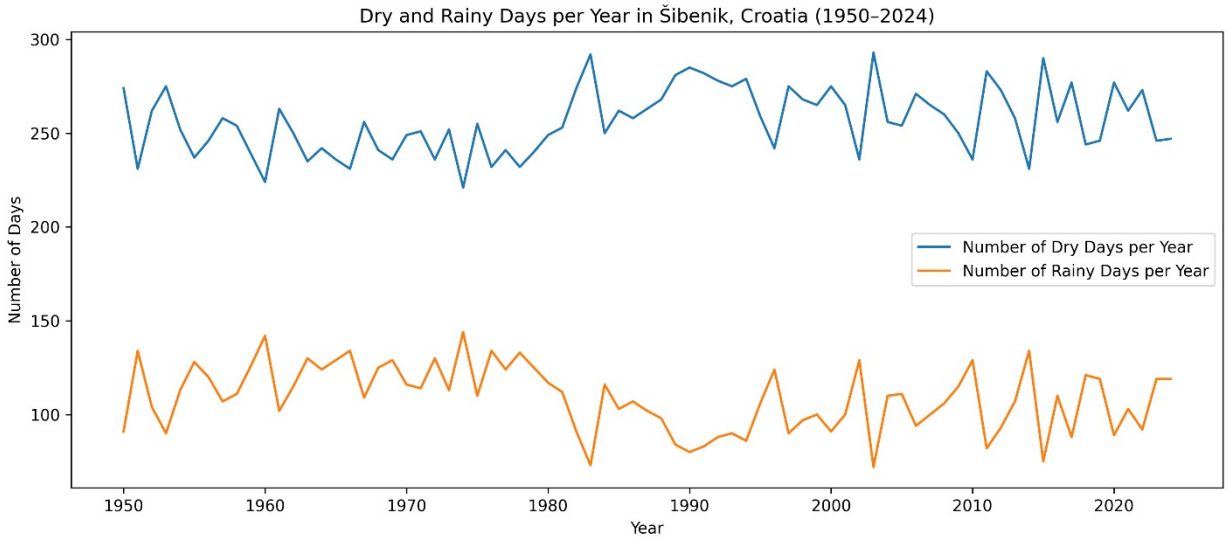


Figura 5.8 Numero di giorni secchi e piovosi all'anno a Sebenico (1950-2024)

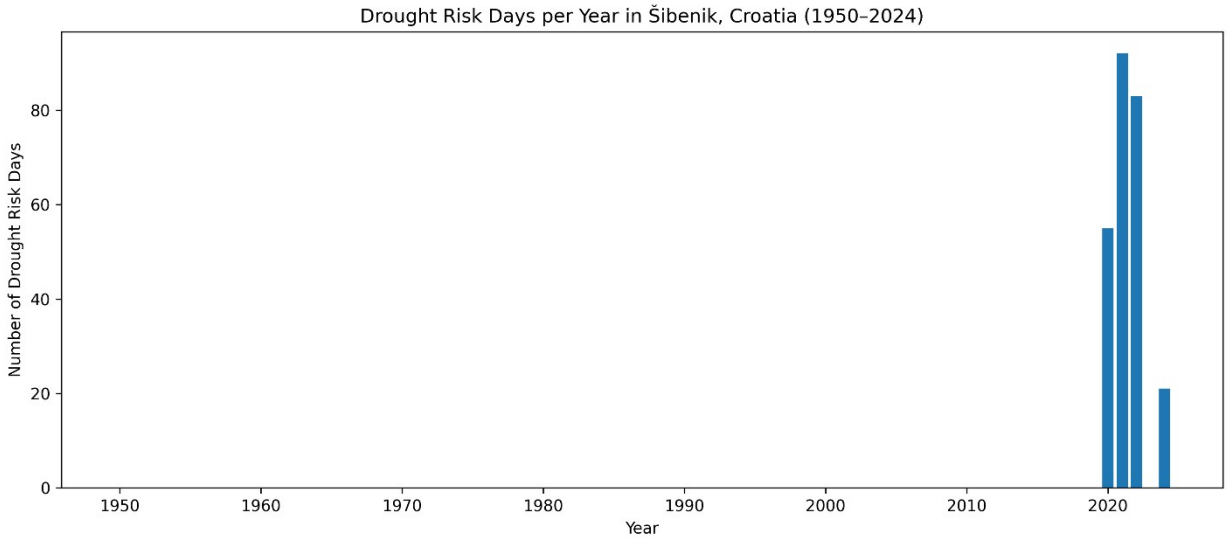


Figura 5.9 Numero di giorni che soddisfano il criterio del rischio di siccità pronunciato a Sebenico (1950-2024)

## Apprendimento automatico non supervisionato per il raggruppamento dei profili climatici annuali

Dopo aver aggregato le misurazioni meteorologiche giornaliere in set di dati annuali di caratteristiche, il metodo di clustering K-means è stato applicato allo spazio di dati. L'analisi è stata eseguita direttamente all'interno del data warehouse di Oracle utilizzando il pacchetto DBMS_DATA_MINING integrato [6], che consente l'integrazione della preparazione, della

modellazione e dell'interpretazione dei risultati dei dati all'interno di un unico framework relazionale.

Un modello di raggruppamento è definito come un modello di cluster (CLUSTERING) applicato alla tabella CLIMATE_YEARLY_FEATURES_SIBENIK, risultato diretto della query di aggregazione SQL descritta nel sottocapitolo precedente. Contiene un record per ogni anno con la corrispondente serie di caratteristiche annuali aggregate, tra cui indicatori termici, idrologici, a secco e eolici.

Il numero di cluster è impostato su 3 nella tabella dei parametri del modello (CLUS_NUM_CLUSTERS = 3). Questa scelta consente una segmentazione equilibrata dei profili climatici annuali in tipi più piccoli e più comprensibili, senza suddividere eccessivamente i risultati.

Il modello K-means è stato addestrato su tutte le funzioni annuali disponibili, con le variabili di input normalizzate e standardizzate automaticamente all'interno del framework Oracle Data Mining. Ciò garantisce che le variabili con intervalli numerici più ampi (ad esempio il numero di giorni o i valori estremi) non dominino il processo di clustering.

Dopo la creazione del modello, ogni anno viene assegnato a un cluster specifico tramite la funzione CLUSTER_ID incorporata. I risultati mostrano che gli anni sono raggruppati in tre gruppi distinti, ciascuno con serie temporali e caratteristiche climatiche diverse, come illustrato nella tabella 1.

Un singolo cluster (ID = 4), che può essere considerato uno scenario climatico stabile (contrassegnato in verde), include il maggior numero di anni ed è caratterizzato da temperature medie più basse, quasi nessuna ondata di calore o di stress secco e il maggior numero di giorni di pioggia. Questo cluster è caratterizzato da anni di climi relativamente stabili e moderati. La Figura 5.10 mostra che questo ammasso domina la maggior parte del periodo osservato, soprattutto nella seconda metà del 20 ° secolo.

Un altro cluster (ID = 5), che rappresenta uno **scenario climatico di transizione** (evidenziato in blu), presenta temperature medie più elevate, ondate di calore più frequenti e giorni più secchi con meno giorni piovosi. Questa combinazione suggerisce un aumento dello stress termico e idrologico, pur non raggiungendo i valori estremi dell'indicatore secco. La figura 5.10 mostra che gli anni in questo cluster risultano sporadici nei periodi precedenti, ma diventano molto più frequenti dopo il 2000.

Il terzo cluster (ID = 3), che rappresenta lo **scenario climatico sfavorevole (estremo)** contrassegnato in arancione, sebbene copra meno anni, è chiaramente caratterizzato da

temperature medie più elevate, da un maggior numero di ondate di calore e da un numero eccezionalmente elevato di giorni con stress arido combinato. Come mostrato nella Figura 5.10, questo cluster si verifica quasi esclusivamente nel periodo di osservazione più recente, indicando uno spostamento qualitativo delle caratteristiche climatiche, più che un cambiamento graduale delle singole variabili.

Tabella 1. Risultati di K-means Clustering of Climate Indicators per Sebenico

| ID cluster | Numero di anni nel cluster | Temperatura media annua (°C) | Sensazione di temperatura media (°C) | Numero medio di ondate di calore all'anno | Media giorni a rischio siccità | Media giorni asciutti all'anno | Numero medio di giorni piovosi all'anno | Velocità massima media del vento (km/h) |
|---|---|---|---|---|---|---|---|---|
| **3** | **4** | **16,5** | **15** | **3,5** | **62,8** | **264,8** | **100,8** | **44,5** |
| 4 | 46 | 14,5 | 12,7 | 0,1 | 0 | 249,4 | 115,8 | 53,7 |
| 5 | 25 | 15,5 | 13,9 | 2,1 | 0 | 268,1 | 97,2 | 51,8 |

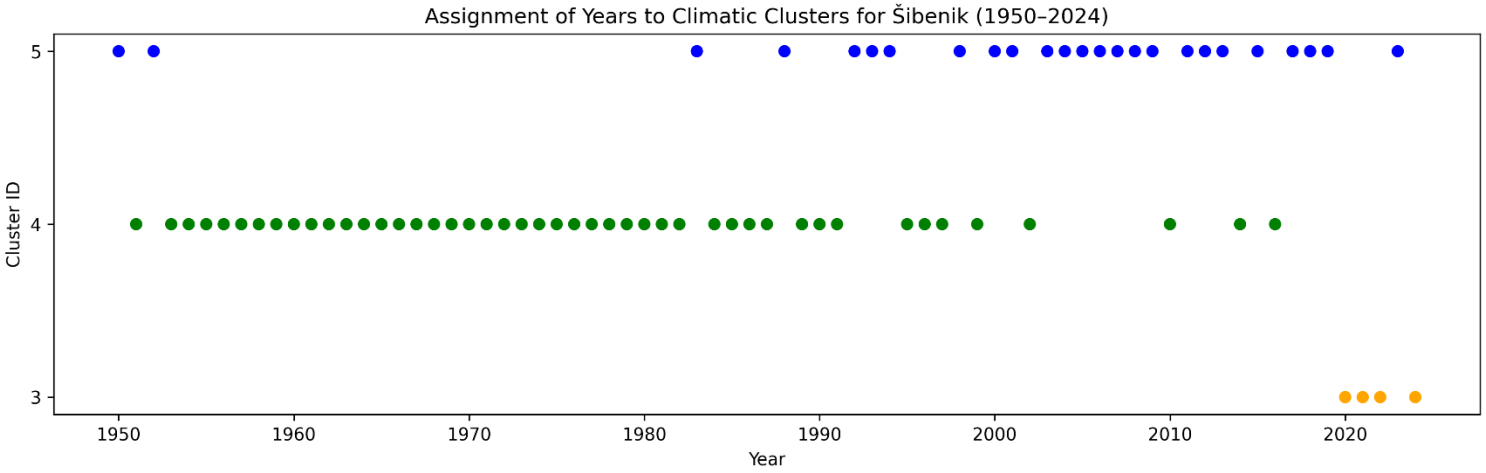


Figura 5.10 Assegnazione di anni ai cluster climatici a Sebenico (1950-2024)

```
-- ======================================================
-- K-means clustering of yearly climate profiles
-- ------------------------------------------------------
-- This script implements unsupervised machine learning
-- directly inside Oracle Autonomous Data Warehouse.
-- The input table CLIMATE_YEARLY_FEATURES_SIBENIK
-- contains one row per year with aggregated climatic
-- indicators derived from daily meteorological data.
--
-- The goal of the analysis is to group years with
-- similar climatic characteristics using the K-means
```

```
-- clustering algorithm.
-- =====================================================

-- -------------------------------------------------------
-- Step 1: Define K-means model settings
-- -------------------------------------------------------
-- Create a settings table used by Oracle Data Mining
-- to configure the clustering algorithm.

CREATE TABLE KMEANS_SETTINGS (
  setting_name  VARCHAR2(30),
  setting_value VARCHAR2(30)
);

-- Specify the number of clusters (K = 3)
INSERT INTO KMEANS_SETTINGS
VALUES ('CLUS_NUM_CLUSTERS', '3');


-- -------------------------------------------------------
-- Step 2: Create the K-means clustering model
-- -------------------------------------------------------
-- The model is trained directly inside the database
-- using the DBMS_DATA_MINING package.
-- Each row represents one year (case),
-- and all remaining columns are used as input features.

BEGIN
  DBMS_DATA_MINING.CREATE_MODEL(
    model_name          => 'CLIMATE_KMEANS_SIBENIK',
    mining_function     => DBMS_DATA_MINING.CLUSTERING,
    data_table_name     => 'CLIMATE_YEARLY_FEATURES_SIBENIK',
    case_id_column_name => 'YEAR',
    settings_table_name => 'KMEANS_SETTINGS'
  );
END;


-- -------------------------------------------------------
-- Step 3: Verify that the model was successfully created
-- -------------------------------------------------------

SELECT model_name, mining_function
FROM user_mining_models
WHERE model_name = 'CLIMATE_KMEANS_SIBENIK';


-- -------------------------------------------------------
-- Step 4: Assign each year to a cluster
-- -------------------------------------------------------
-- The CLUSTER_ID function returns the cluster
-- to which each yearly climate profile belongs.
```

```
SELECT
  year,
  CLUSTER_ID(CLIMATE_KMEANS_SIBENIK USING *) AS cluster_id
FROM CLIMATE_YEARLY_FEATURES_SIBENIK
ORDER BY year;


-- ------------------------------------------------------
-- Step 5: Profile clusters using average feature values
-- ------------------------------------------------------
-- This query summarizes the climatic characteristics
-- of each cluster by calculating average values
-- of selected indicators.

SELECT
  CLUSTER_ID(CLIMATE_KMEANS_SIBENIK USING *) AS cluster_id,
  COUNT(*) AS years_in_cluster,

  -- Temperature characteristics
  ROUND(AVG(avg_year_temp), 1) AS avg_temp,
  ROUND(AVG(avg_apparent_temp), 1) AS avg_apparent_temp,

  -- Heat stress indicators
  ROUND(AVG(heatwave_count), 1) AS avg_heatwaves,

  -- Drought-related indicators
  ROUND(AVG(drought_risk_days), 1) AS avg_drought_days,
  ROUND(AVG(dry_days), 1) AS avg_dry_days,

  -- Precipitation characteristics
  ROUND(AVG(rainy_days), 1) AS avg_rainy_days,

  -- Wind extremes
  ROUND(AVG(max_wind_speed_km_h), 1) AS avg_max_wind

FROM CLIMATE_YEARLY_FEATURES_SIBENIK
GROUP BY CLUSTER_ID(CLIMATE_KMEANS_SIBENIK USING *)
ORDER BY cluster_id;
```

## Conferma di modelli climatici identificati utilizzando dati recenti

I risultati del caso di studio nell'area di Sebenico (Croazia) mostrano che i cambiamenti climatici sono evidenti non solo attraverso aumenti graduali della temperatura, ma anche attraverso cambiamenti nella struttura degli anni climatici e combinazioni avverse più frequenti di condizioni termiche e idrologiche. L'analisi visiva degli indicatori annuali rivela un aumento della frequenza e della durata delle ondate di calore, nonché il verificarsi di anni con stress secco combinato significativo.

I risultati dell'analisi del cluster utilizzando K-means hanno indicato che, dal 2020, è emerso un tipo distinto di anno climatico nell'area osservata, caratterizzato dallo stress termico e secco più pronunciato. Questo cluster si distingue per alte temperature medie e massime, ondate di calore più frequenti e più lunghe, e periodi prolungati di stress secco combinato; nel tempo, appare quasi esclusivamente nel periodo di analisi più recente.

L'analisi dei dati meteorologici quasi completi per il 2025, presentati nella tabella 2 (conclusa il 12 dicembre 2025), mostra che le caratteristiche di quest'anno si allineano strettamente con il modello precedentemente identificato. Sono state registrate quattro onde termiche di durata media superiore ai 13 giorni, con una temperatura massima giornaliera estremamente elevata di 38,4 °C e più di 100 giorni di stress secco combinato, mentre il numero di giorni di pioggia è diminuito. Questa combinazione di indicatori termici e idrologici corrisponde al profilo del cluster più estremo identificato nell'analisi dei dati storici.

Ciò conferma che l'analisi applicata dei K-means non è soltanto una descrizione retrospettiva delle condizioni passate, ma anche uno strumento per rilevare i cambiamenti strutturali nel comportamento climatico del sistema. Pertanto, l'anno 2025 non è un estremo isolato, ma parte di un modello iniziato nei primi mesi del 2020s, sottolineando l'importanza di questo approccio per l'identificazione precoce di scenari avversi e per la valutazione del rischio ambientale.

Tabella 2 Indicatori climatici chiave per il 2025 (Šibenik)

| Indicatore | Val. |
|---|---|
| Numero di ondate di calore | 4 |
| Durata media ondata termica (giorni) | 13,5 |
| Temperatura massima giornaliera (°C) | 38,4 |
| Temperatura media annua dell'aria (°C) | 17 |
| Sensazione di temperatura media (°C) | 15,3 |
| Giorni asciutti | 252 |
| Numero di giorni di pioggia | 94 |
| Numero di giorni con stress secco combinato | 101 |
| Numero di giorni con aumento dell'evapotraspirazione | 128 |
| Velocità massima del vento (km/h) | 48,4 |

## Implicazioni per la riduzione del rischio ambientale basata su scenari

Nell'area di Šibenik, i cluster climatici identificati fungono da scenari basati su prove che illustrano diversi livelli di rischio ambientale. Gli anni con condizioni climatiche relativamente stabili fungono da scenario di base. Tuttavia, un cluster è caratterizzato da un significativo stress da calore e da siccità, che rappresenta uno scenario sfavorevole. La crescente frequenza di

questi anni negli ultimi decenni, insieme a dati quasi completi per il 2025, indica che queste gravi condizioni stanno diventando una caratteristica regolare del clima locale, anziché un evento raro.

In pratica, l'aumento della frequenza degli anni con tratti climatici sfavorevoli sottolinea la necessità di una migliore gestione delle risorse idriche, soprattutto durante i lunghi periodi di siccità e durante la stagione turistica, quando la domanda di acqua aumenta. La crescente frequenza e durata delle ondate di calore evidenziano l'importanza delle azioni preventive nella salute pubblica, nell'adattamento urbano e nell'uso dell'energia, in particolare nelle aree affollate e nei mesi estivi. Inoltre, la comprensione del passaggio da tipi di clima più stabili a quelli più rischiosi può aiutare a sviluppare sistemi di allarme precoce locali e piani di adattamento ai cambiamenti climatici.

## Estensione dell'analisi alle città aggiuntive

Per verificare l'applicabilità dell'approccio proposto, sono state condotte ulteriori analisi su tre città in diverse regioni geografiche e climatiche: Roma (Italia), Skopje (Macedonia del Nord) e Kiev (Ucraina).

Per ogni città è stato utilizzato lo stesso processo analitico di Šibenik, compresa la selezione delle variabili meteorologiche, l'aggregazione delle misurazioni giornaliere ai dati annuali e l'applicazione dell'algoritmo K-means. Tuttavia, il modello è stato addestrato separatamente su dati specifici relativi alla posizione. Ciò garantisce che i cluster identificati rappresentino i modelli climatici locali e le relazioni tra gli indicatori all'interno di ogni città.

I risultati dell'analisi dei cluster per Roma, Skopje e Kiev sono presentati nelle relative tabelle (3-5), che mostrano le medie degli indicatori climatici chiave per cluster, nonché nelle figure (5.11-5.13), che illustrano la distribuzione degli anni climatici per cluster nel periodo 1950-2024. Sebbene la metodologia sia coerente in tutti i casi (come in precedenza per Šibenik), i cluster ottenuti differiscono per caratteristiche e tendenze temporali, confermando che i modelli climatici e i rischi variano tra i contesti regionali.

Tabella 3 Risultati di K-means Clustering di indicatori climatici per Roma

| ID cluster | Numero di anni nel cluster | Temperatura media annua (°C) | Sensazione di temperatura media (°C) | Numero medio di ondate di calore all'anno | Media giorni a rischio siccità | Media giorni asciutti all'anno | Numero medio di giorni piovosi all'anno | Velocità massima media del vento (km/h) |
|---|---|---|---|---|---|---|---|---|
| 3 | 20 | 16 | 15,3 | 3 | 40,3 | 268,7 | 96,6 | 35,3 |
| 4 | 21 | 15,4 | 14,8 | 2,8 | 15,1 | 255,5 | 109,8 | 34,4 |
| 5 | 34 | 14,7 | 14 | 1,2 | 5 | 250,1 | 115,1 | 34,4 |

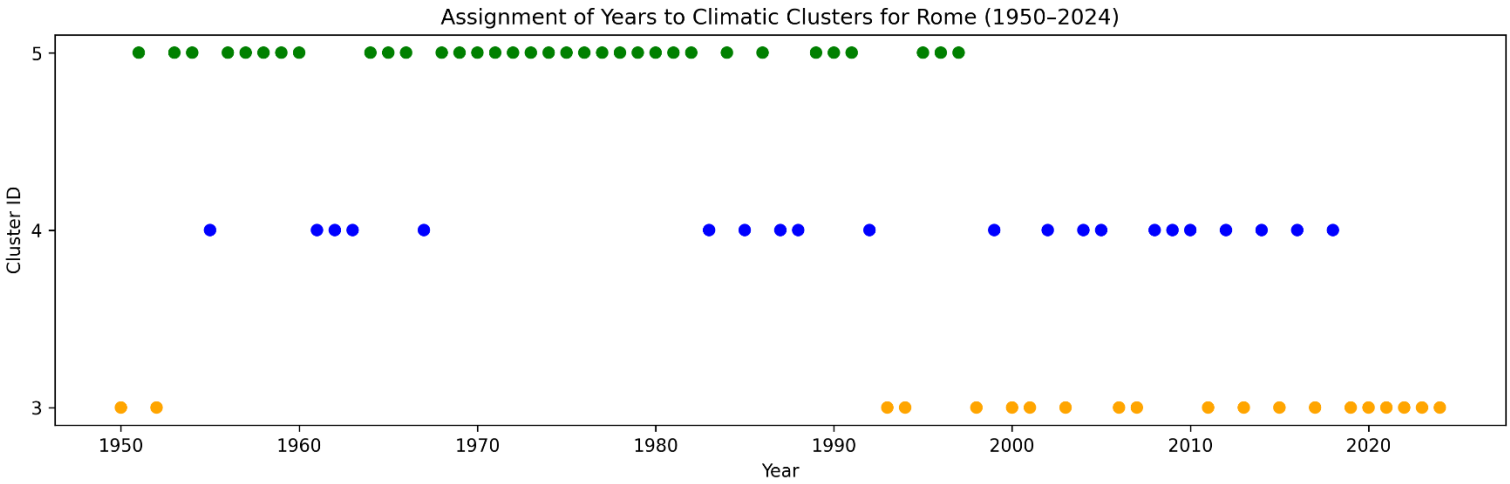


Figura 5.10. Assegnazione di anni ai cluster climatici a Roma (1950-2024)

Tabella 4 Risultati di K-means Clustering di indicatori climatici per Skopje

| ID cluster | Numero di anni nel cluster | Temperatura media annua (°C) | Sensazione di temperatura media (°C) | Numero medio di ondate di calore all'anno | Media giorni a rischio siccità | Media giorni asciutti all'anno | Numero medio di giorni piovosi all'anno | Velocità massima media del vento (km/h) |
|---|---|---|---|---|---|---|---|---|
| 2 | 45 | 12,6 | 11,1 | 2,6 | 36,6 | 269,8 | 95,4 | 26,7 |
| 4 | 3 | 13,1 | 11,4 | 2,3 | 66,7 | 301,3 | 64 | 31 |
| 5 | 27 | 14,1 | 12,6 | 4,6 | 78,2 | 285,1 | 80,1 | 26,2 |

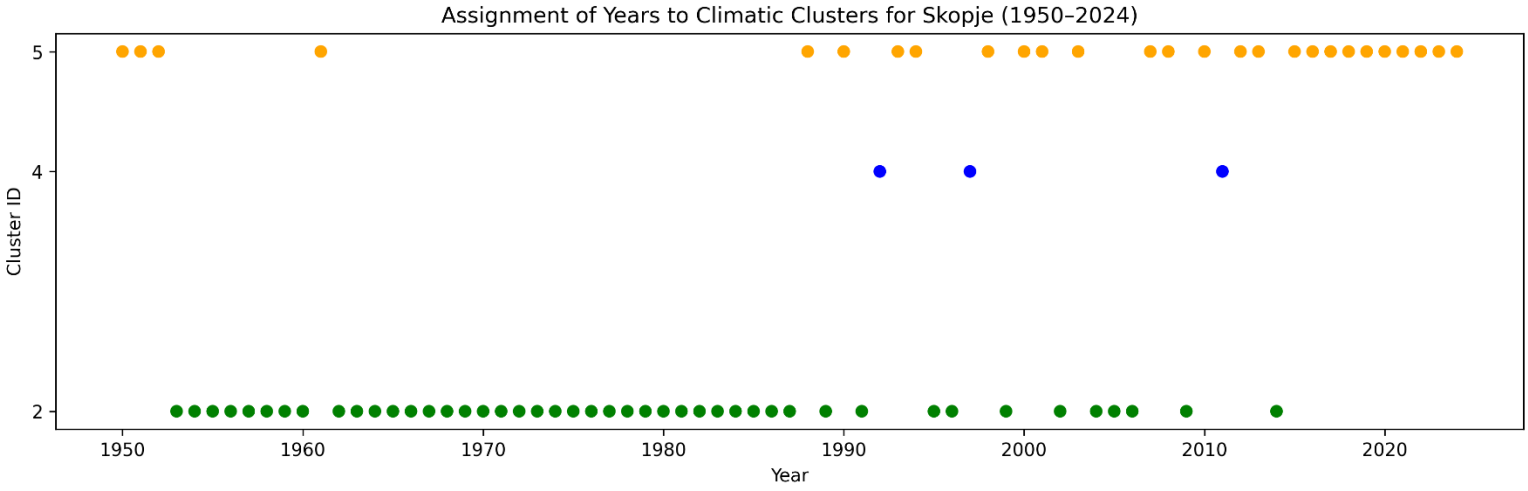


Figura 5.11 Assegnazione di anni ai cluster climatici di Skopje (1950-2024)

Tabella 5 Risultati di K-means Clustering di indicatori climatici per Kiev

| ID cluster | Numero di anni nel cluster | Temperatura media annua (°C) | Sensazione di temperatura media (°C) | Numero medio di ondate di calore all'anno | Media giorni a rischio siccità | Media giorni asciutti all'anno | Numero medio di giorni piovosi all'anno | Velocità massima media del vento (km/h) |
|---|---|---|---|---|---|---|---|---|
| 3 | 2 | 7 | 4,1 | 0 | 0 | 248 | 117,5 | 52,5 |
| 4 | 63 | 7,7 | 4,8 | 0 | 0 | 243,2 | 122 | 39,6 |
| 5 | 10 | 8,9 | 6,1 | 1,2 | 0 | 255,6 | 109,6 | 37,1 |

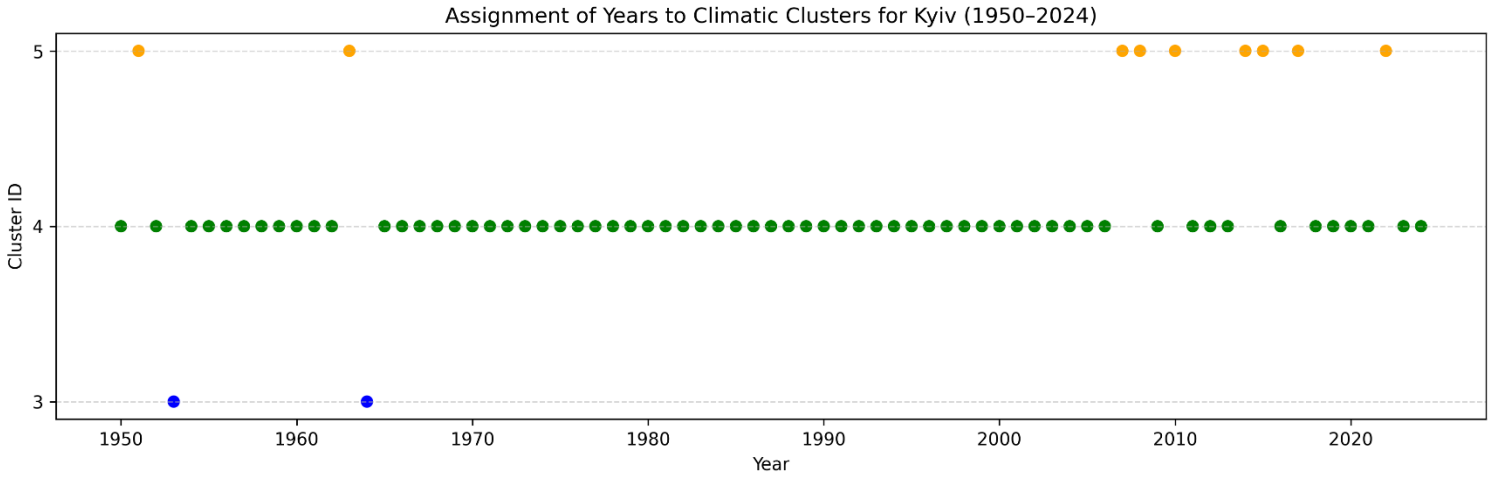


Figura 5.11 Assegnazione di anni ai cluster climatici a Kiev (1950-2024)

L'analisi di altre città mostra che a Roma, Skopje e Kiev si è registrato un recente spostamento verso anni con condizioni climatiche estreme. Tuttavia, questi modelli differiscono. In tutte e tre le città, la frequenza degli anni con temperature più elevate e con maggiore stress termico è aumentata. Al contrario, i modelli climatici più stabili si verificano più spesso nella parte più antica della serie di dati.

È particolarmente interessante notare che, dai primi anni 2000 e ancor più dopo il 2010, i cluster più estremi sono diventati più comuni nelle città continentali come Skopje e Kiev, indicando la diffusione dello stress termico oltre le tipiche regioni del Mediterraneo. Sebbene l'intensità e le combinazioni di condizioni estreme differiscano da quelle delle città costiere, i risultati suggeriscono un cambiamento strutturale nei modelli climatici anche in queste aree.

Questi risultati confermano che i modelli identificati non sono soltanto anomalie locali, ma fanno parte di una tendenza più ampia verso il cambiamento climatico, con scenari estremi che

emergono gradualmente in diverse zone climatiche, evidenziando la necessità di metodi analitici localmente adattati ma applicati in modo coerente.

## Attività di esercizio

- Scarica dati meteo giornalieri storici per la posizione selezionata utilizzando (ad esempio, API Open-Meteo)
- Aggrega i dati meteorologici giornalieri in indicatori climatici annuali (ad esempio, temperatura media, numero di giorni molto caldi, numero di giorni asciutti).
- Stabilire profili climatici annuali basati su indicatori selezionati.
- Applicare l'algoritmo K-means per raggruppare gli anni climatici in base alle loro caratteristiche.
- Assegnare ogni anno al proprio cluster e visualizzare i risultati in formato tabulare o grafico.
- Interpretare i cluster ottenuti come diversi scenari climatici: stabili, transitori e sfavorevoli.
- Confronta i dati più recenti con i cluster identificati in precedenza e valuta a quale tipo di anno climatico appartengono.

## Valutazione

Qual è l'obiettivo principale dell'applicazione dell'algoritmo K-means nel caso di studio presentato?

A. Previsione numerica temperatura futura
B. Raggruppamento degli anni climatici in base alle loro somiglianze
C. Calcolo delle medie stagionali delle variabili meteorologiche
D. Rilevamento di singoli giorni estremi

Qual è lo scopo di aggregare i dati meteorologici giornalieri a un livello annuale?

A. Riduci il volume dei dati senza perdere informazioni
B. Più facile individuare modelli a lungo termine e tipi di clima
C. Rimuovi stagionalità dai dati
D. Prepara i dati di apprendimento monitorati

Quale serie di indicatori descrive meglio lo scenario climatico estremo identificato nel caso di studio?

A. Alte temperature, giorni piovosi elevati, vento basso
B. Temperature moderate e ondate di calore rare

C. Temperature elevate, ondate di calore frequenti e un gran numero di giorni con stress secco combinato
D. Grande quantità di precipitazioni e estremi a breve termine

Perché il cluster identificato dopo il 2020, nel caso della città di Sebenico, è considerato un indicatore dei cambiamenti strutturali nelle condizioni climatiche?

A. Perché si basa su un solo anno estremo
B. Perché include nuove combinazioni di stress termico e idrologico
C. Perché appare quasi esclusivamente nel periodo recente
D. Perché le soglie lo predefiniscono

Quale affermazione descrive meglio il ruolo dell'algoritmo K-means nella previsione dei rischi ambientali?

A. K-means prevede direttamente le condizioni climatiche future come scenari
B. K-means identifica i pattern tipici che possono essere interpretati come scenari
C. K-means utilizza l'analisi meteorologica classica come scenari
D. K-means è utilizzato per le previsioni a breve termine

## Origini aggiuntive

[1] G. James, D. Witten, T. Hastie, and R. Tibshirani, An Introduction to Statistical Learning: With Applications in R, 2nd ed. New York, NY, USA: Springer, 2021 [Online]. Available: https://hastie.su.domains/ISLR2/ISLRv2_website.pdf, Accessed: Dec. 15, 2025

[2] Lindholm, N. Wahlström, F. Lindsten, and T. B. Schön, Machine Learning – A First Course for Engineers and Scientists. Cambridge, U.K.: Cambridge University Press, 2022 [Online]. Available: http://smlbook.org/book/sml-book-draft-latest.pdf, Accessed: Dec. 15, 2025

[3] O. Simeone, A brief introduction to machine learning for engineers, arXiv preprint, arXiv:1709.02840, 2017. [Online]. Available: https://arxiv.org/pdf/1709.02840.pdf, Accessed: Dec. 15, 2025

[4] Historical Weather API Documentation, 2024. Available: https://open-meteo.com/en/docs/historical-weather-api, Accessed: Dec. 15, 2025.

[5] Oracle, Autonomous AI Lakehouse. [Online]. Available: https://www.oracle.com/autonomous-database/autonomous-ai-lakehouse, Accessed: Dec. 15, 2025.

[6] Oracle, DBMS_DATA_MINING Package Reference, Oracle Database 19c Documentation. [Online]. Available: https://docs.oracle.com/en/database/oracle/oracle-database/19/arpls/DBMS_DATA_MINING.html, Accessed: Dec. 15, 2025.