



Energy Learning Report

Presentato da: Fabio Catello Ponticelli [0512116210] e Giovanni Ruocco [0512116171]

Sommario

[1. Introduzione](#)

[2. Descrizione del problema](#)

[3. Dataset utilizzato](#)

[3.1 Caratteristiche del dataset](#)

[3.5 Data Engineering](#)

[3.5.1 Gestione degli outlier](#)

[3.5.2 Dati parzialmente assenti](#)

[3.6 Feature Engineering](#)

[4 Modello](#)

[5. Test Set](#)

[6. Analisi delle prestazioni](#)

[7. Considerazioni finali](#)

[7.1 Struttura del Progetto](#)

[7.2 Limiti e Possibili Evoluzioni](#)

[7.3 Conclusioni](#)



1. Introduzione

L'obiettivo di questo studio è sviluppare un modello predittivo basato sulla tendenza temporale mensile del consumo di elettricità in Italia. Analizzando i dati storici relativi ai consumi elettrici su base mensile, miriamo a identificare pattern e fluttuazioni stagionali che possano guidare previsioni accurate per i consumi futuri. Questo approccio semplificato si concentra esclusivamente sul comportamento temporale del consumo, garantendoci stime affidabili. Il nostro modello sarà utile per ottimizzare la gestione della domanda energetica e pianificare strategie di approvvigionamento elettrico, contribuendo a una gestione più efficiente delle risorse.

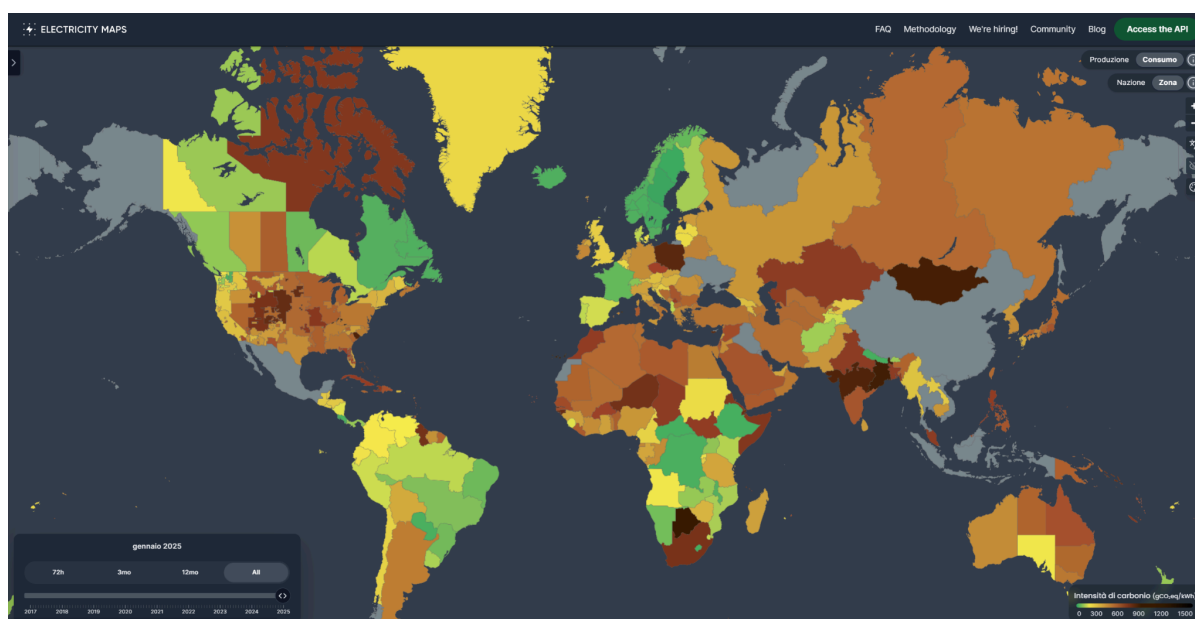
2. Descrizione del problema

Il consumo di elettricità in Italia presenta una forte dipendenza dalla stagionalità, con variazioni mensili che influenzano direttamente la domanda energetica. Le fluttuazioni dei consumi seguono schemi ricorrenti legati ai cambiamenti stagionali, come l'aumento della domanda nei periodi di maggiore caldo o freddo. Questa tendenza temporale è un elemento cruciale.

3. Dataset utilizzato

Il dataset da noi utilizzato è quello reso disponibile dal sito [Electricity Maps](https://electricitymaps.com/), il quale raccoglie e divide in tempo reale dati sulla produzione elettrica divulgati dai principali enti nazionali.

Il sito dà un'ulteriore opzione dividendo alcuni paesi in zone in base alle loro peculiari situazioni energetiche. Queste caratteristiche sono prettamente di natura geografica per via della posizione di risorse naturali e di import/export nazionale. Prenderemo però l'Italia come nazione, in quanto ci risparmierebbe lavoro che già è stato fatto da Electricity Maps.



Screenshot dal sito Electricity Maps

3.1 Caratteristiche del dataset

timestampKey: Data e ora di riferimento della rilevazione, in formato ISO 8601.

totalConsumption: Quantità totale di energia consumata nella regione in **MWh**.

La seguente struttura JSON rappresenta la entry relativa all'Italia nel mese di dicembre 2023:

```
{
  "_id": {
    "$oid": "67a4d1c024237056ebd71b1a"
  },
  "zoneKey": "IT",
  "timestampKey": "2023-01-01T00:00:00Z",
  "co2intensity": 398.32,
```



```
"co2intensityProduction": 466.84,  
"completenessPercentage": 100,  
"estimatedPercentage": 0,  
"fossilFuelRatio": 0.5726,  
"fossilFuelRatioProduction": 0.6798,  
"maxDischarge": 0,  
"maxExport": 0,  
"maxImport": 0,  
"maxProduction": 8875741.44,  
"maxStorage": 9180.96,  
"maxStorageCapacity": 0,  
"measuredPercentage": 100,  
"renewableRatio": 0.3428,  
"renewableRatioProduction": 0.3202,  
"totalCo2Consumption": {  
  "$numberLong": "9563039271552"  
},  
"totalCo2Production": 9033328118390.4,  
"totalConsumption": 24008433.6,  
"totalDischarge": 166336.08,  
"totalExport": 311788.08,  
"totalImport": 4979465.52,  
"totalProduction": 19349944.56,  
"totalStorage": 175517.04,  
"dischargeCo2IntensitiesBattery": 288.18,  
"dischargeCo2IntensitiesHydro": 288.18,  
"priceCurrency": null,  
"priceValue": null,  
"productionBiomass": 491568.24,  
"productionCo2IntensitiesBiomass": 230,  
"productionCo2IntensitiesCoal": 1061.28,  
"productionCo2IntensitiesGas": 535.98,  
"productionCo2IntensitiesGeothermal": 38,  
"productionCo2IntensitiesHydro": 10.7,  
"productionCo2IntensitiesNuclear": 5.13,  
"productionCo2IntensitiesOil": 1104.73,  
"productionCo2IntensitiesSolar": 26.26527778,  
"productionCo2IntensitiesUnknown": 700,  
"productionCo2IntensitiesWind": 12.62,  
"productionCoal": 2309532.24,  
"productionGas": 8875741.44,  
"productionGeothermal": 458891.76,  
"productionHydro": 2019387.12,  
"productionNuclear": 0,  
"productionOil": 391812.72,  
"productionSolar": 834604.32,  
"productionUnknown": 1698812.4,  
"productionWind": 2269601.76,  
"isFinestGranularity": false
```

}

Il motivo principale per cui i dati sono stati trasferiti in **MongoDB** è la sua capacità di gestire dati semi-strutturati in modo più flessibile rispetto a un database relazionale. Questo approccio offre diversi vantaggi:

- **Scalabilità e velocità di accesso** ai dati, facilitando l'elaborazione di grandi volumi di informazioni.
- **Flessibilità nella gestione di dati eterogenei**, poiché nuovi campi possono essere aggiunti senza modificare la struttura del database.
- **Facile integrazione con pipeline di analisi dati e strumenti di visualizzazione**, consentendo aggiornamenti e query più efficienti.

3.5 Data Engineering

3.5.1 Gestione degli outlier

Nel marzo 2020, il consumo di elettricità in Italia ha registrato un calo significativo rispetto ai trend storici. Questo fenomeno è probabilmente attribuibile alle misure di lockdown adottate per contenere la pandemia di COVID-19, che hanno ridotto drasticamente le attività industriali, commerciali e dei trasporti. Nonostante questa anomalia, abbiamo scelto di mantenere tale mensilità nel dataset, ritenendo che rappresenti un evento rilevante piuttosto che un errore o un outlier da rimuovere. La presenza di questo dato potrebbe influenzare la previsione, ma è importante considerarlo per comprendere meglio l'impatto di eventi straordinari sulla domanda di energia.

3.5.2 Dati parzialmente assenti

Nel gennaio 2025, Electricity Maps ha stimato il **47%** della produzione elettrica in Italia. Tuttavia, questa stima rappresentava un'anomalia rispetto ai dati storici e ha creato un problema per la qualità della nostra analisi. Per evitare che questo dato distorcesse le previsioni, abbiamo deciso di rimuovere quella mensilità dal dataset.

3.6 Feature Engineering

Nel contesto di questo progetto, il processo di **feature engineering** si è concentrato principalmente sulla variabile temporale mensile, che rappresenta l'andamento **stagionale** del consumo elettrico. I dati storici del consumo di elettricità, espressi in terawattora (**TWh**) per ogni mese, sono stati utilizzati per identificare tendenze e pattern ricorrenti legati alle variazioni stagionali. Non sono state considerate altre feature esterne, come la composizione della produzione elettrica, per semplificare il modello e concentrarsi esclusivamente sull'andamento del consumo. Tuttavia, particolari mensilità, come quella di marzo 2020, in cui si è verificato un calo anomalo dovuto alla pandemia di **COVID-19**, sono state mantenute nel dataset per preservare l'integrità dei dati storici. D'altro canto, la mensilità di gennaio 2025, in cui è stata stimata una percentuale particolarmente elevata di produzione elettrica, è stata rimossa dal dataset poiché tale valore si discostava notevolmente dai dati storici, rischiando di distorcere le previsioni. L'attenzione è stata quindi rivolta a garantire che il dataset contenesse solo i dati più **coerenti** e **rappresentativi** per la previsione del consumo elettrico.

4 Modello

Abbiamo scelto **Prophet** per questo progetto di previsione del consumo elettrico in Italia principalmente per la sua capacità di gestire serie temporali con forti componenti stagionali e trend non lineari, caratteristiche tipiche dei dati di consumo energetico. Prophet, sviluppato da Facebook, è particolarmente adatto a situazioni in cui i dati presentano trend di lunga durata, effetti stagionali annuali e settimanali, e altre anomalie o eventi eccezionali, come quelli derivanti da situazioni pandemiche o crisi economiche. Inoltre, la sua facilità d'uso e la possibilità di includere variabili esterne (regressori), come ad esempio la percentuale di energia prodotta da fonti rinnovabili, lo rendono un potente strumento per analisi predittive. La flessibilità di Prophet nella gestione dei cambiamenti improvvisi nei dati, insieme alla sua robustezza rispetto agli outlier, lo ha reso una scelta ideale per trattare i dati energetici, che sono spesso influenzati da fattori stagionali e imprevedibili.

5. Test Set

Per la creazione del test set, abbiamo selezionato il **20%** delle mensilità più recenti del dataset, utilizzando un rapporto di divisione **80:20**. Questa scelta si basa sulla natura del nostro obiettivo, che è quello di prevedere il consumo di elettricità per i mesi futuri, partendo dai dati storici. Poiché i dati più recenti sono quelli più rilevanti per catturare eventuali cambiamenti nelle tendenze dei consumi, l'uso delle mensilità più recenti come test set ci consente di valutare la performance del modello su periodi temporali più simili a quelli di previsione. Inoltre, mantenere il **80%** dei dati per l'addestramento ci permette di avere un ampio campione di dati storici su cui il modello può imparare, senza compromettere l'affidabilità dei risultati. In questo modo, il test set riflette in modo accurato la capacità del modello di generalizzare a situazioni future, mantenendo la coerenza temporale dei dati a disposizione.

6. Analisi delle prestazioni

Abbiamo utilizzato i valori di MAE, MSE e MAPE, essi sono metriche utilizzate per valutare la **performance** di un modello di previsione, confrontando i valori previsti con quelli reali.

MAE: 0.46

MSE: 0.39

MAPE: 1.93%

Il **MAE** misura la **media** delle **differenze assolute** tra i valori previsti e quelli reali. Indica quanto, in media, il modello si discosta dal valore effettivo. Un MAE basso suggerisce che il modello è accurato.

Il **MSE** calcola la **media degli errori al quadrato** tra i valori reali e quelli previsti. Essendo elevato al quadrato, penalizza maggiormente gli errori più grandi rispetto al MAE, quindi è più sensibile agli outlier. Un MSE basso indica che gli errori sono in generale piccoli.

Il **MAPE** esprime l'errore come una **percentuale** rispetto ai valori reali. È una metrica molto utile per comparare il modello su dataset di diversa scala, poiché fornisce un'indicazione relativa degli errori. Un MAPE basso indica che il modello è più preciso.

I risultati delle metriche di errore (MAE, MSE e MAPE) indicano una buona performance del modello di previsione, con valori che suggeriscono un'accuratezza complessiva soddisfacente, ma ci potrebbero essere miglioramenti soprattutto per ridurre gli errori più grandi.

7. Considerazioni finali

7.1 Struttura del Progetto

L'approccio adottato prevede la raccolta dei dati storici, la loro elaborazione e la costruzione di un modello predittivo per stimare i consumi futuri. Il modello Prophet è stato scelto per la sua capacità di gestire componenti stagionali e tendenze a lungo termine, risultando particolarmente adatto all'analisi dei dati energetici, caratterizzati da ciclicità annuale e fluttuazioni dipendenti da fattori esterni.

Per valutare la qualità delle previsioni, sono stati calcolati gli errori MAE, MSE e MAPE, che offrono un'indicazione quantitativa dell'accuratezza del modello. I risultati mostrano che il modello è in grado di catturare le dinamiche generali del consumo energetico, pur con alcune limitazioni dovute a fluttuazioni impreviste o eventi eccezionali.

7.2 Limiti e Possibili Evoluzioni

Attualmente, il progetto si configura come un processo batch offline, il che significa che l'aggiornamento dei dati e delle previsioni avviene periodicamente, in base alla disponibilità delle informazioni storiche. Questo approccio è stato adottato per garantire una corretta elaborazione e correzione dei dati prima dell'input nel modello.

Tuttavia, una possibile evoluzione del progetto potrebbe consistere nell'integrazione di un sistema online e in tempo reale, a condizione che sia disponibile un flusso di dati affidabile e continuo, ad esempio da fonti come Electricity Maps. Un'architettura basata su data streaming permetterebbe di aggiornare le previsioni dinamicamente, migliorando l'accuratezza delle analisi e fornendo informazioni utili in tempo reale agli stakeholder del settore energetico.

7.3 Conclusioni

Il progetto sviluppato offre una base solida per l'analisi e la previsione del consumo energetico in Italia, permettendo di anticipare eventuali variazioni nei trend di consumo e supportare la pianificazione energetica nazionale. L'adozione di modelli più sofisticati e l'integrazione con flussi di dati in tempo reale potrebbero rappresentare i prossimi passi per migliorare ulteriormente l'affidabilità e l'utilità delle previsioni fornite.