

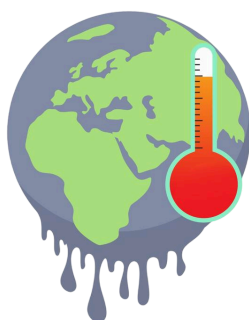


UNIVERSITÀ
DEGLI STUDI DI BARI
ALDO MORO

DIPARTIMENTO DI
INFORMATICA

ECO-MACHINE LEARNING

Analisi dei Cambiamenti Climatici e Sviluppo di Interventi Sostenibili
attraverso l'Apprendimento Automatico



caso di studio per il corso di:

Ingegneria della conoscenza

a.a. 2023/2024

Prof. Nicola Fanizzi

CdS in Informatica

3° anno, 1° semestre, 6 CFU

Link GitHub: [qui](#)

Studenti:

Emilio Cives, Matricola 755578, e.cives@studenti.uniba.it

Vincenzo Ficco, Matricola 758340, v.ficco5@studenti.uniba.it

INDICE

INDICE.....	1
INTRODUZIONE.....	1
STRUTTURA DEL PROGETTO.....	2
ELENCO ARGOMENTI D'INTERESSE.....	3
REQUISITI.....	4
DESCRIZIONE DEL DOMINIO E DATASET.....	4
PREPROCESSING DEI DATI.....	5
Climate Change: Earth Surface Temperature Data.....	5
Social Economic Data.....	7
FEATURES.....	8
ANALISI DESCRITTIVA DEL DATASET.....	9
ANDAMENTO MEDIO DEI CAMBIAMENTI CLIMATICI DAL 1850 AL 2015.....	11
PREVALENZA E VARIABILITÀ DEGLI IMPATTI CLIMATICI NEL MONDO.....	12
APPRENDIMENTO NON SUPERVISIONATO.....	13
CLUSTERING AGGLOMERATIVO.....	14
KMEANS CLUSTERING.....	17
CONCLUSIONI.....	20
APPRENDIMENTO SUPERVISIONATO.....	22
ANALISI DEI RISULTATI.....	24
APPROCCIO ENSEMBLE.....	26
RANDOM FOREST VS MLP (MULTI LAYER PERCEPTRON).....	27
OTTIMIZZAZIONE DEL RANDOM FOREST.....	29
CONCLUSIONI.....	31
ONTOLOGIA.....	32
ENVIRONMENT ONTOLOGY (ENVO).....	33
METODOLOGIA.....	34
CONCLUSIONI FINALI.....	36
IDEE PER ESTENSIONI FUTURE.....	36

INTRODUZIONE

Negli ultimi anni, la crisi climatica è diventata un tema centrale nel dibattito pubblico e scientifico. Questo progetto si focalizza sui principali fenomeni legati ai cambiamenti climatici, come l'aumento delle temperature, la desertificazione, l'innalzamento del livello del mare e la perdita di biodiversità.

L'obiettivo è di sviluppare interventi economico-sociali mirati, attraverso l'analisi di dati climatici e l'uso di tecniche di apprendimento automatico.

Attraverso l'analisi dei dati climatici e l'uso di tecniche di apprendimento automatico, questo progetto mira a comprendere meglio i cambiamenti climatici e a sviluppare interventi sostenibili e mirati.

La creazione di un'ontologia ambientale facilita la condivisione dei dati e l'interoperabilità tra diversi sistemi, contribuendo così a un approccio più integrato e efficace nella lotta contro i cambiamenti climatici.

STRUTTURA DEL PROGETTO

- **Analisi dei Dati Storici Climatici:** Prima di procedere con le analisi, esamineremo i dati storici relativi ai cambiamenti climatici per comprenderne l'andamento e le tendenze. Questa fase preliminare è fondamentale per ottenere una visione globale della situazione attuale e delle dinamiche passate, grazie soprattutto all'utilizzo di grafici e visualizzazioni.
- **Clustering delle Regioni:** Abbiamo applicato tecniche di apprendimento non supervisionato per identificare profili di regioni simili in termini di impatto climatico, come aumento delle temperature e frequenza di eventi climatici estremi. Questo approccio ci ha permesso di raggruppare le regioni in cluster, facilitando così la proposta di interventi personalizzati e mirati alle specifiche esigenze di ciascun gruppo.
- **Predizione degli Impatti Ambientali:** Parallelamente, abbiamo utilizzato tecniche di apprendimento supervisionato per predire gli impatti ambientali in specifiche regioni. Questo include la predizione dei cambiamenti nei modelli meteorologici, l'innalzamento del livello del mare e la perdita di biodiversità. Predire questi impatti ci consente di comprendere meglio l'entità del problema e di pianificare interventi più efficaci.
- **Sviluppo di un'Ontologia Ambientale:** Infine, abbiamo sviluppato un'ontologia per organizzare i nostri dati e risultati riguardanti i cambiamenti climatici, gli impatti ambientali e i gruppi di intervento. L'obiettivo principale è facilitare la condivisione, il riutilizzo e l'interoperabilità dei dati tra diversi sistemi e applicazioni. Questo approccio consente l'uso di inferenze e ragionamenti automatici, migliorando significativamente l'analisi dei dati. L'integrazione della nostra ontologia con le **ontologie ambientali esistenti** ci permette di utilizzare definizioni e relazioni standardizzate già esistenti. Questo garantisce che i nostri

dati siano compatibili con altre fonti di dati e strumenti che adoperano la stessa ontologia, aumentando così la validità e l'utilità dei dati.

ELENCO ARGOMENTI D'INTERESSE

Descrizione del Dominio e Dataset

- Origine dei dati
- Descrizione dei dataset utilizzati
- Fonti e validità dei dati

Preprocessing dei Dati

- Pulizia dei dati
- Unione dei dataset
- Eliminazione di colonne vuote e rinominazione di colonne
- Standardizzazione e normalizzazione dei dati

Analisi Descrittiva

- Matrice di correlazione
- Andamento medio dei cambiamenti climatici dal 1850 al 2015
- Prevalenza e variabilità degli impatti climatici tra le diverse regioni
- Visualizzazioni grafiche e interpretazioni

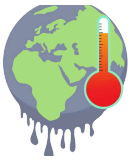
Apprendimento Non Supervisionato

- Modelli utilizzati (Agglomerative Clustering e KMeans Clustering)
- Integrazione dei dati socio-economici (GDP)
- Risultati e analisi dei cluster
- Descrizione dei cluster ottenuti
- Interpretazione dei risultati
- Gruppi d'intervento ottenuti

Apprendimento Supervisionato

- Obiettivi e preparazione dei dati
- Modelli predittivi utilizzati
- Regressione lineare
- Regressione polinomiale
- Random Forest
- Multi-Layer Perceptron (MLP)
- Ottimizzazione dei modelli e metriche di valutazione (RMSE, MAE, R^2)
- Risultati e analisi comparativa
- Previsioni future degli impatti climatici per le diverse regioni
- Visualizzazione e interpretazione delle previsioni

Ontologia



- Creazione dell'ontologia
- Integrazione con le ontologie ambientali esistenti
- Mappatura degli impatti climatici agli URI corrispondenti
- Creazione e serializzazione del grafo RDF
- Esempi di query SPARQL
- Verifica della struttura dell'ontologia

Conclusioni

- Limiti dello studio e considerazioni per estensioni future

REQUISITI

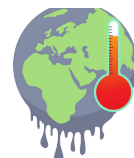
Il codice Python è stato creato per essere eseguito direttamente all'interno dell'ambiente di sviluppo integrato (IDE). L'esecuzione è semplificata grazie a menu interattivi che consentono di selezionare diverse opzioni.

Sono state utilizzate varie librerie per i vari passi:

- **Preprocessing**
 - pandas
 - sklearn
- **Visualizzazione dei grafici**
 - plotly
 - matplotlib
 - pandas
 - numpy
 - seaborn
 - geopandas
 - shapely
- **Apprendimento supervisionato**
 - sklearn
- **Apprendimento non supervisionato**
 - sklearn
 - Kneelocator
- **Ontologie**
 - rdflib

DESCRIZIONE DEL DOMINIO E DATASET

Il dataset utilizzato per le nostre analisi proviene da [Kaggle](#) e include informazioni dettagliate sui cambiamenti climatici e sugli impatti ambientali, fornendo una base solida per le nostre analisi.



In seguito ad una ricerca più approfondita, abbiamo constatato che i dati provengono da diverse fonti autorevoli come la NASA, il National Oceanic and Atmospheric Administration (NOAA) e altre organizzazioni internazionali che monitorano i cambiamenti climatici. Questo dataset è utilizzato per numerosi rapporti e studi sui cambiamenti climatici, come evidenziato dalla documentazione disponibile su Kaggle.

Inoltre questo dataset è stato rifatto da una compilazione più recente messa insieme dal [Berkeley Earth](#), che è affiliato al Lawrence Berkeley National Laboratory. Il Berkeley Earth Surface Temperature Study combina 1,6 miliardi di report sulla temperatura da 16 archivi preesistenti.

Questo dataset ha ricoperto un ruolo principale per le nostre analisi, ma è stato considerato anche un altro dataset messo a disposizione dalla World Bank Open Data: “[World Development Indicators.csv](#)”, poi rinominato “SocialEconomicData.csv”, il cui uso sarà specificato nella sezione dedicata all’apprendimento non supervisionato.

PREPROCESSING DEI DATI

La fase introduttiva consiste nel leggere il dataset e comprenderne la struttura, focalizzandosi sulla risoluzione di alcune problematiche legate alle celle vuote e l’individuazione di correlazioni tra la colonna target e le altre feature.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import plotly
import plotly.express as px
```

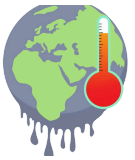
Climate Change: Earth Surface Temperature Data

Inizialmente i dati a nostra disposizione si dividevano in 5 dataset distinti

1. GlobalLandTemperaturesByCity

Temperature globali terrestri per città

dt	Entity	Code	Average Temperature	Average Temperature Uncertainty	Latitude	Longitude
Data di rilevazione AAAA-MM-GG	Stringa: nome della nazione, paese o area	Stringa di dimensione 3 che identifica le nazioni	Float: temperatura media globale in gradi	Float: l'intervallo di incertezza per la	Stringa: latitudine di rilevazione	Stringa: longitudine di rilevazione



			Celsius	media		
--	--	--	---------	-------	--	--

2. GlobalLandTemperaturesByCountry

Temperatura media globale della terra per paese

dt	Average Temperature	Average Temperature Uncertainty	Country
Data di rilevazione AAAA-MM-GG	Float: temperatura media globale in gradi Celsius	Float: l'intervallo di incertezza per la media	Stringa: paese di rilevazione

3. GlobalLandTemperaturesByMajorCity

Temperature globali terrestri per città principali

dt	Average Temperature	Average Temperature Uncertainty	City	Country	Latitude	Longitude
Data di rilevazione AAAA-MM-GG	Float: temperatura media globale in gradi Celsius	Float: l'intervallo di incertezza per la media	Stringa: città di rilevazione	Stringa: paese di rilevazione	Stringa: latitudine di rilevazione	Stringa: longitudine di rilevazione

4. GlobalLandTemperaturesByState

Temperatura media globale terrestre per stato

dt	Average Temperature	Average Temperature Uncertainty	State	Country
Data di rilevazione AAAA-MM-GG	Float: temperatura media globale in gradi Celsius	Float: l'intervallo di incertezza per la media	Stringa: stato di rilevazione	Stringa: paese di rilevazione

5. GlobalTemperatures

Temperature globali terrestri e oceaniche

dt	Entity	code	Avera	Avera	Land	Land	Land	Land	Ocean	Ocean
----	--------	------	-------	-------	------	------	------	------	-------	-------

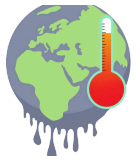
			ge Temp eratur e	ge Temp eratur e Uncer tainty	MaxTe mpera ture	MaxTe mpera tureU ncerta inty	MinTe mpera ture	MinTe mpera tureU ncerta inty	Avera geTe mpera ture	Avera geTe mpera tureU ncerta inty
Data di rilevazi one AAAA- MM-G G	String a: nome della nazion e, paese o area	String a di dimen sione 3 che identifi ca le nazion i	Float: temper atura media global e della terra in gradi Celsiu s	Float: l'interv allo di incerte zza per la media	Float: temper atura media massi ma global e della terra in gradi Celsiu s	Float: l'interv allo di incerte zza per la media	Float: temper atura media minim a global e della terra in gradi Celsiu s	Float: l'interv allo di incerte zza per la media	Float: temper atura media global e degli oceani in gradi Celsiu s	Float: l'interv allo di incerte zza per la media

Tra i 5 database disponibili, per rilevanza e semplicità sono stati considerati il numero 1 e 5: questi due dataset contengono dati sulla temperatura media globale, temperatura media massima globale, temperatura media minima globale e temperatura media globale degli Oceani.

Durante la fase di preprocessing, sono state eseguite diverse operazioni significative per garantire che i dati fossero adeguatamente puliti e pronti per l'analisi successiva; tra cui:

- **pulizia dei dati:** abbiamo rimosso le righe con valori mancanti o non validi per garantire la qualità e l'integrità dei dati; con questa operazione sono state eliminate le righe del periodo dal 1750 al 1849.
- **eliminazione di colonne vuote e rinominazione di colonne:** abbiamo eliminato le colonne non pertinenti e rinominato le colonne per maggiore chiarezza e coerenza (tipo la colonna dt in data)
- **eliminazione di colonne inutili:** abbiamo eliminato le colonne che riguardano la percentuale d'incertezza in quanto non erano utili nel nostro studio.
- **standardizzazione e normalizzazione dei dati:** abbiamo standardizzato e normalizzato i dati delle temperature per facilitare le analisi successive
- **merge tra i due database considerati,** ottenendo un unico database di nome climateChange. L'obiettivo di questa operazione è stato quello di ottenere un set di dati più completo e informativo per l'analisi successiva, garantendo al contempo che non ci fossero duplicati o informazioni ridondanti.

Con questi passaggi, ci siamo assicurati di avere un dataset pulito e strutturato, pronto per le analisi successive sull'andamento dei cambiamenti climatici e per l'applicazione di tecniche di apprendimento automatico.



Social Economic Data

Questo dataset è strutturato in questo modo

Series Name	Series Code	Country Name	Country Code	1990 [YR1990], 2000 [YR2000], 2014 [YR2014], 2015 [YR2015], 2016 [YR2016], 2017 [YR2017], 2018 [YR2018], 2019 [YR2019], 2020 [YR2020], 2021 [YR2021], 2022 [YR2022], 2023 [YR2023]
Stringa: nome della serie di dati	Stringa: codice della serie di dati	Stringa: nome della nazione a cui sono associati i valori	Stringa: codice corrispondente alla nazione	Double: valore del PIL per lo specifico anno

Nella fase di preprocessing

- è stata rinominata la colonna “Country Code” in “Code” per renderlo coerente con il dataset precedente
- sono state eliminate le colonne “2022 [YR2022]”, “2023 [YR2023]” in quanto vuote

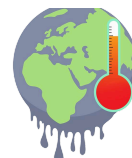
In tutte queste situazioni, abbiamo utilizzato Python e librerie come Pandas per gestire e selezionare i dati, assicurandoci che solo le informazioni rilevanti fossero mantenute per l'analisi successiva.

FEATURES

Da questa fase di preprocessing, abbiamo ottenuto le seguenti features:

CLIMATECHANGE

date	entity	code	Latitude	Longitude	LandAverage Temperature	LandMax Temperature	LandMin Temperature	OceanAverage Temperature
data di rilevazio	nome della	identifica tore	latitudin e di	longitudi ne di	temperat ura	temperat ura	temperat ura	temperat ura



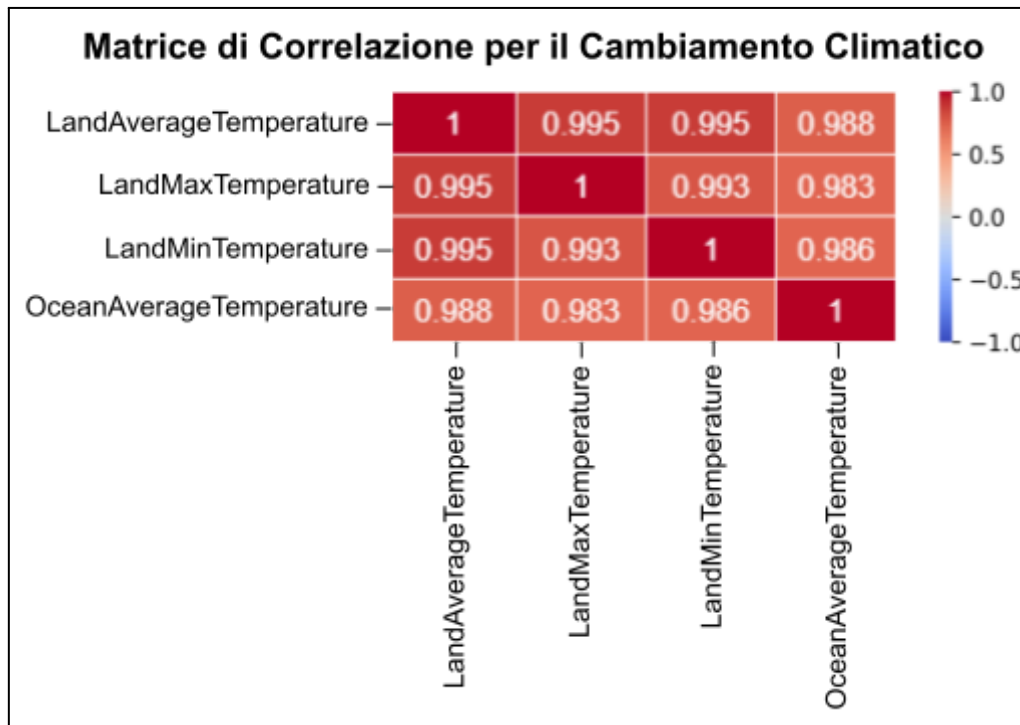
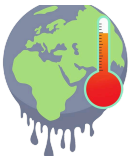
ne	nazione, paese o area	delle nazioni	rilevazio ne	rilevazio ne	media globale della terra	media massim a globale della terra	media minima globale della terra	media globale degli oceani
Data	Stringa	Stringa	Stringa	Stringa	Float	Float	Float	Float

SOCIALECONOMICDATA

Series Name	Series Code	Country Name	Code	1990 [YR1990], 2000 [YR2000], 2014 [YR2014], 2015 [YR2015], 2016 [YR2016], 2017 [YR2017], 2018 [YR2018], 2019 [YR2019], 2020 [YR2020], 2021 [YR2021]
nome della serie di dati	codice della serie di dati	nome della nazione a cui sono associati i valori	Stringa: codice corrispondente alla nazione	Double: valore del PIL per lo specifico anno
Stringa	Stringa	Stringa	Stringa	Double

ANALISI DESCRITTIVA DEL DATASET

Abbiamo esaminato la matrice di correlazione per ottenere informazioni sulle relazioni tra i diversi fenomeni climatici e variabili ambientali, facilitando una migliore comprensione delle loro interconnessioni e influenze reciproche.



Questa matrice mostra i coefficienti di correlazione tra le temperature medie, massime, minime terrestri e la temperatura media dell'oceano.

I valori sono tutti molto vicini a 1, indicando una forte correlazione positiva tra le variabili considerate e quindi forti relazioni lineari tra queste variabili climatiche.

La tabella è stata calcolata utilizzando i dati forniti e considerando solo le variabili scelte dopo aver eliminato le righe vuote e inutilizzabili nella fase di preprocessing: è stata calcolata utilizzando 1,992 campioni dopo la pulizia dei dati.

Le **variabili climatiche** mostrano alcune correlazioni significative:

- Tra le correlazioni positive, la temperatura media della terra e la temperatura massima della terra (0.995) sono particolarmente rilevanti.
- Per quanto riguarda le correlazioni negative, non ci sono correlazioni negative significative tra queste variabili.

La **temperatura massima** della terra presenta anch'essa correlazioni interessanti:

- Le correlazioni positive includono la temperatura minima della terra (0.993) e la temperatura media degli oceani (0.983).
- Non ci sono correlazioni negative significative.

Per quanto riguarda la **temperatura minima** della terra:

- Si osservano forti correlazioni positive con la temperatura media della terra (0.995) e la temperatura media degli oceani (0.986).
- Non ci sono correlazioni negative significative.

La **temperatura media degli oceani** mostra:

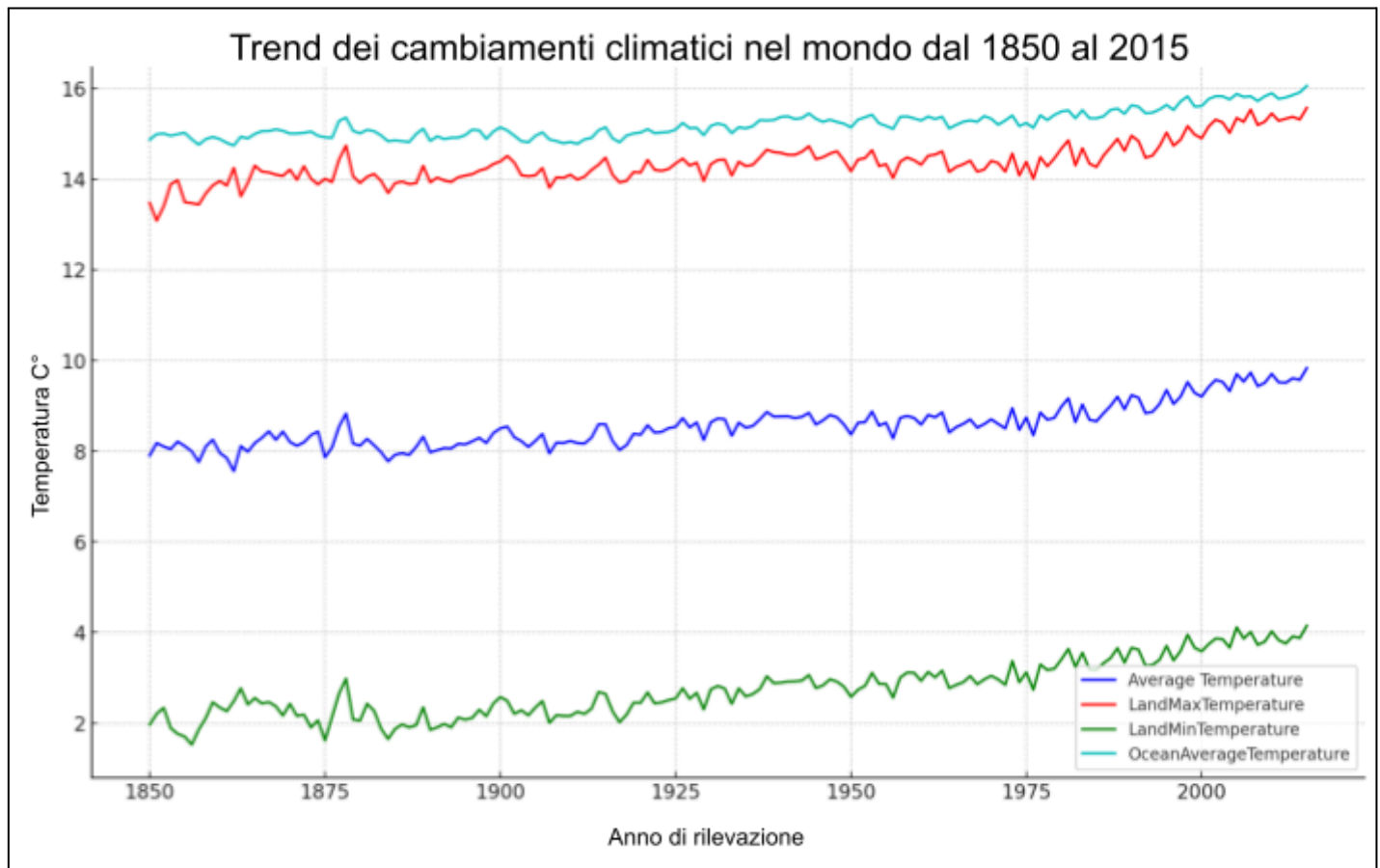
- Correlazioni positive significative con la temperatura media della terra (0.988) e la temperatura massima della terra (0.983).
- Non si registrano correlazioni negative significative.

Le correlazioni positive significative tra la temperatura media della terra e la temperatura massima e minima della terra, così come la temperatura media degli oceani, suggeriscono che queste variabili climatiche tendono a variare insieme. Inoltre, la co-occorrenza forte tra le diverse temperature della terra indica che questi fenomeni climatici spesso si verificano insieme.

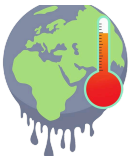
ANDAMENTO MEDIO DEI CAMBIAMENTI CLIMATICI DAL 1850 AL 2015

Abbiamo realizzato un grafico per analizzare l'andamento dei cambiamenti climatici nel periodo compreso tra il 1850 e il 2015.

Nonostante il database contenesse dati dal 1750, alcune colonne del periodo 1750-1850 erano vuote e quindi sono state eliminate durante la fase di preprocessing.



Dall'analisi del grafico possiamo dedurre che le temperature medie annuali presentano una tendenza al rialzo nel tempo, con le seguenti osservazioni specifiche:



- La **Average Temperature** (Temperatura Media) mostra un aumento costante nel tempo, indicando un trend di riscaldamento globale.
- La **LandMaxTemperature** (Temperatura Massima Terrestre) ha una tendenza simile, con un aumento significativo nel corso degli anni e soprattutto degli ultimi decenni
- La **LandMinTemperature** (Temperatura Minima Terrestre) evidenzia una crescita, anche se a un ritmo leggermente inferiore rispetto alle temperature massime.
- La **OceanAverageTemperature** (Temperatura Media degli Oceani) ha una tendenza al rialzo meno accentuata rispetto alle temperature terrestri, ma comunque evidente.

Le tendenze indicano che le temperature terrestri e oceaniche sono aumentate negli ultimi decenni, con le temperature terrestri che mostrano una maggiore variabilità.

La LandMaxTemperature e la Average Temperature sono generalmente più alte rispetto alla LandMinTemperature e alla OceanAverageTemperature.

Le temperature oceaniche aumentano, ma a un ritmo meno rapido rispetto alle temperature terrestri, riflettendo la maggiore capacità termica degli oceani che riscaldano e raffreddano più lentamente rispetto alla terra.

Questi dati suggeriscono che il riscaldamento globale ha un impatto significativo sulle temperature medie annuali, con un aumento generale che potrebbe avere implicazioni importanti per il clima e gli ecosistemi globali.

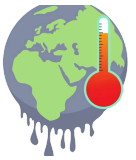
PREVALENZA E VARIABILITÀ DEGLI IMPATTI CLIMATICI NEL MONDO

La prevalenza media degli impatti climatici, come **l'aumento delle temperature**, si attesta su una crescita globale dell' $1,2^{\circ}\text{C}$ negli ultimi decenni, con una **deviazione standard di $0,3^{\circ}\text{C}$** . Questo dato indica che, in media, la temperatura della superficie terrestre è **aumentata di $1,2^{\circ}\text{C}$** rispetto ai livelli preindustriali.

La bassa deviazione standard suggerisce una relativa uniformità nella tendenza al riscaldamento tra le diverse regioni, indicando una certa omogeneità nei fattori di rischio globali come le emissioni di gas serra.

La prevalenza dell'aumento della temperatura è rimasta relativamente **stabile in alcune regioni**, ma ha mostrato variazioni significative in altre, con picchi di riscaldamento più marcati in aree come l'Artico e l'Africa settentrionale. Questo aumento delle temperature potrebbe essere attribuito a una combinazione di fattori naturali e antropogenici, inclusi **l'uso di combustibili fossili e le pratiche agricole**.

La **temperatura massima terrestre** mostra una variabilità leggermente superiore, con un aumento medio di $1,1^{\circ}\text{C}$ e una deviazione standard di $0,3^{\circ}\text{C}$. Questo suggerisce che alcune



regioni possono aver sperimentato periodi di caldo estremo più pronunciati, probabilmente a causa di condizioni climatiche locali e fenomeni meteorologici

La **temperatura minima terrestre**, invece, ha registrato un aumento medio di $0,7^{\circ}\text{C}$, con una deviazione standard di $0,2^{\circ}\text{C}$. Questo indica che anche le temperature notturne o invernali stanno aumentando, contribuendo a una riduzione delle escursioni termiche giornaliere e stagionali.

Per quanto riguarda gli **oceani**, la temperatura media ha mostrato un aumento medio di $0,6^{\circ}\text{C}$, con una **deviazione standard di $0,15^{\circ}\text{C}$** . Questa variazione più uniforme rispetto alle terre emerse è indicativa della grande capacità termica degli oceani, che tendono ad assorbire e rilasciare calore in modo più graduale.

Questi dati indicano una tendenza generale al riscaldamento globale, con variazioni regionali che possono essere attribuite a diversi fattori, tra cui la latitudine, la vicinanza agli oceani, e l'urbanizzazione.

L'uso di modelli di regressione lineare per analizzare i trend nel tempo ha permesso di identificare le regioni con il maggiore incremento degli impatti climatici. Questo approccio ha facilitato il calcolo della pendenza della retta di regressione, rappresentando il tasso di cambiamento annuale per ogni impatto climatico.

Ecco una versione modificata per adattarla al contesto dell'analisi dei cambiamenti climatici:

La **pendenza**, ottenuta tramite la funzione di regressione lineare `np.polyfit`, ci fornisce due parametri: la pendenza stessa (coefficiente angolare) e l'intercetta.

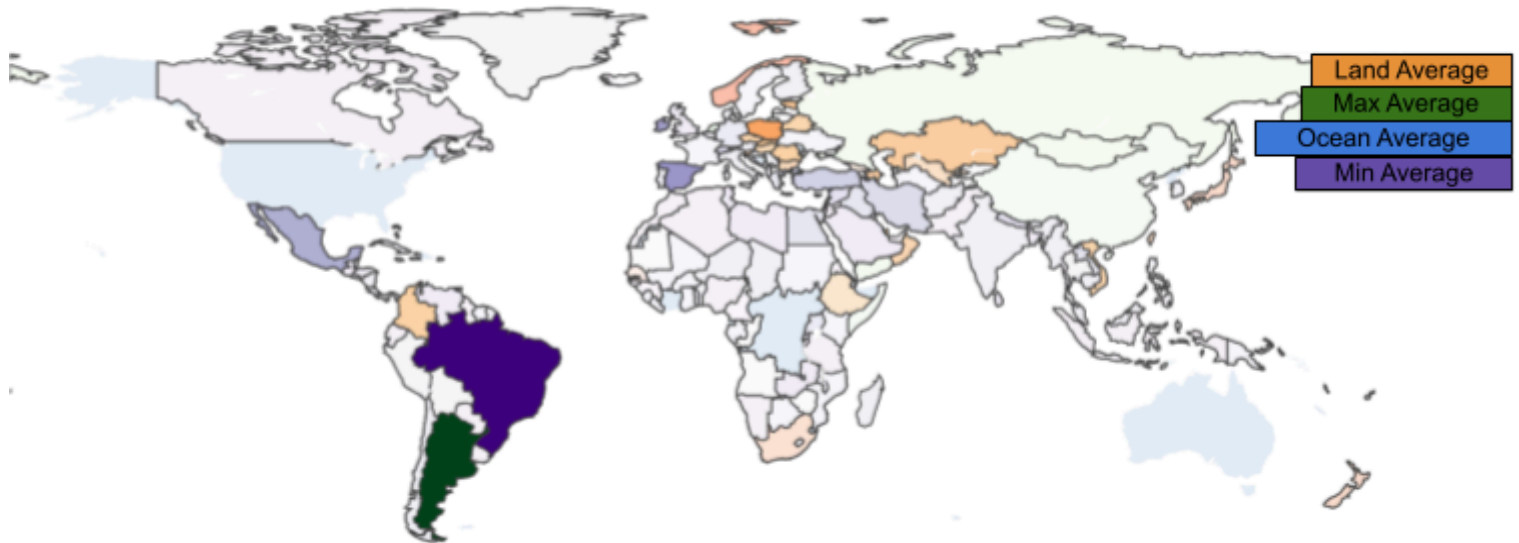
Tuttavia, ai fini della nostra analisi, abbiamo considerato solo la pendenza poiché essa indica il tasso di incremento o decremento della temperatura nel tempo.

Una pendenza positiva implica un aumento della temperatura, mentre una negativa suggerisce una diminuzione.

L'**intercetta** fornirebbe il valore di base della temperatura in un anno iniziale ipotetico, utile per comprendere il punto di partenza della tendenza nel tempo, ma non essenziale per l'analisi delle variazioni nel contesto dei cambiamenti climatici.

Dopo aver calcolato la pendenza per ciascuna variabile di temperatura in ogni regione, abbiamo filtrato i dati per selezionare solo quelle regioni che mostrano una pendenza positiva, indicando un incremento delle temperature nel tempo.

Nazioni con incidenza di cambiamenti climatici in crescita



La visualizzazione dei risultati è stata realizzata attraverso una mappa, che rappresenta graficamente i paesi con tendenze crescenti, utilizzando le librerie di python geopandas e matplotlib

APPRENDIMENTO NON SUPERVISIONATO

L'obiettivo di questa analisi è identificare cluster distinti di regioni con profili climatici simili, per facilitare lo sviluppo di interventi specifici e politiche sostenibili adeguate a ciascun cluster.

Sono stati utilizzati due metodi di clustering: **Agglomerative Clustering** e **KMeans Clustering**, per confrontare la loro efficacia e determinare quale metodo fornisca una migliore separazione dei cluster climatici.

Per garantire un confronto equo e accurato tra le regioni, i dati relativi alle temperature (media, massima e minima terrestre, e media degli oceani) sono stati normalizzati utilizzando la tecnica di **StandardScaler**. Questa normalizzazione è cruciale per assicurare che le diverse variabili climatiche siano comparabili e che non ci siano distorsioni dovute a scale di misura diverse.

Durante l'analisi, è stata integrata l'informazione socio-economica, come il **Prodotto Interno Lordo** (PIL), dal dataset "World Development Indicators.csv" della World Data Banks. Il PIL, che misura il valore monetario totale di tutti i beni e servizi prodotti all'interno di un paese in un periodo specifico, di solito un anno, costituisce un indicatore fondamentale per valutare la performance economica delle nazioni.

Questa misura ha permesso di approfondire l'analisi, fornendo un quadro più completo delle influenze economiche sul cambiamento climatico. In particolare, ha consentito di esplorare come

le variazioni economiche possano influenzare le capacità di adattamento e risposta delle diverse regioni ai cambiamenti climatici, arricchendo la comprensione delle dinamiche tra variabili climatiche ed economiche.

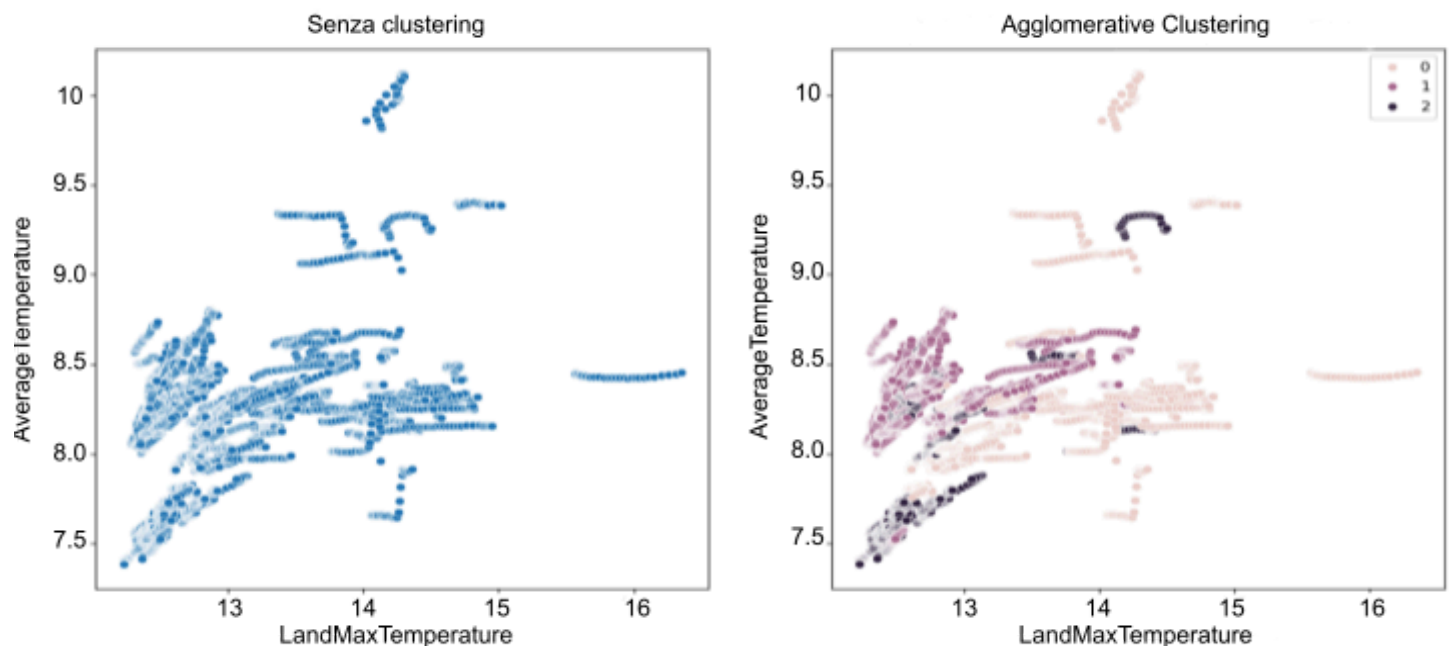
CLUSTERING AGGLOMERATIVO

Il clustering agglomerativo è una tecnica di clustering gerarchico che costruisce un albero (dendrogramma) di cluster unendo iterativamente i punti dati simili. Esistono due principali approcci per il clustering gerarchico: **agglomerativo** e **divisivo**. Nel clustering agglomerativo, si inizia con ogni punto dato come un cluster separato e si procede unendo i cluster più simili fino a ottenere un singolo cluster che contiene tutti i punti dati.

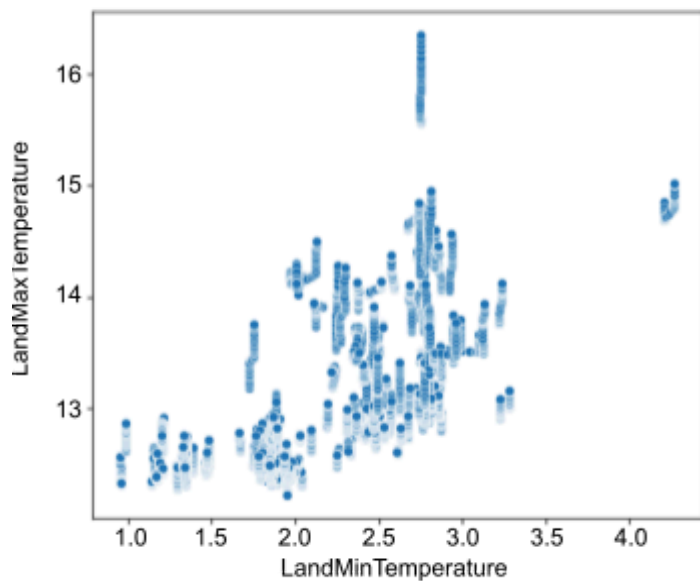
Il clustering agglomerativo è stato eseguito utilizzando il metodo di collegamento Ward (``linkage="ward"``), che minimizza la varianza all'interno dei cluster durante la fusione, per ottenere cluster distinti. I risultati hanno mostrato che:

- il **Cluster 0** include principalmente regioni dell'Europa occidentale e dell'America Latina,
- il **Cluster 1** comprende molte regioni dell'Asia, dell'Europa orientale e del Pacifico,
- il **Cluster 2** è composto principalmente da regioni africane.

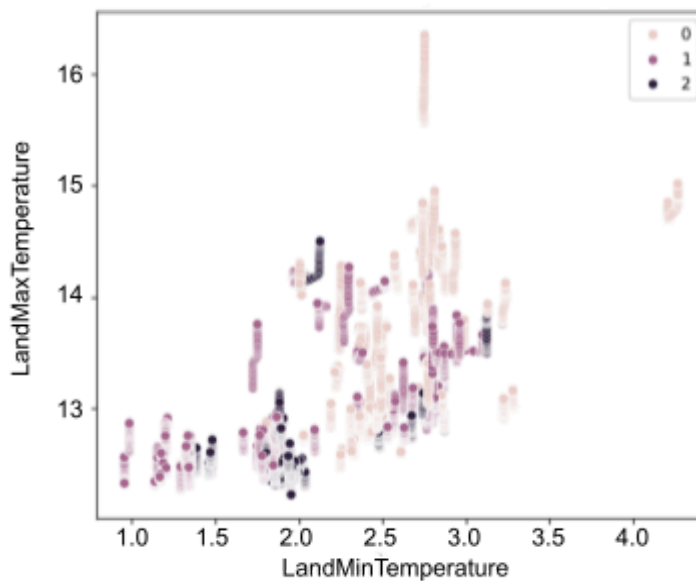
Questo metodo ha prodotto un coefficiente di silhouette di 0.388, che misura la qualità della separazione dei cluster. Sebbene questo coefficiente indichi una discreta separazione, suggerisce anche la presenza di alcuni punti che potrebbero essere vicini ai confini tra i cluster.



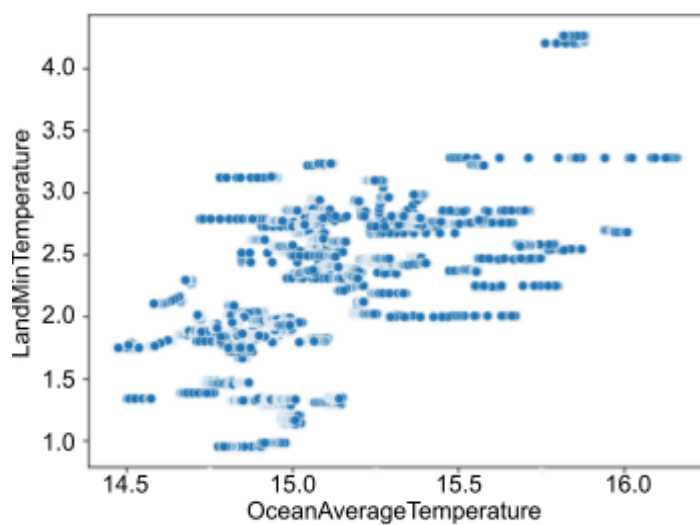
Senza clustering



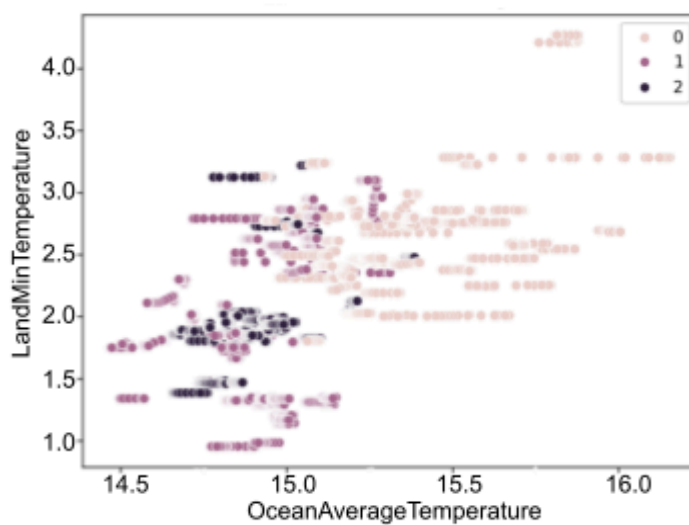
Agglomerative Clustering

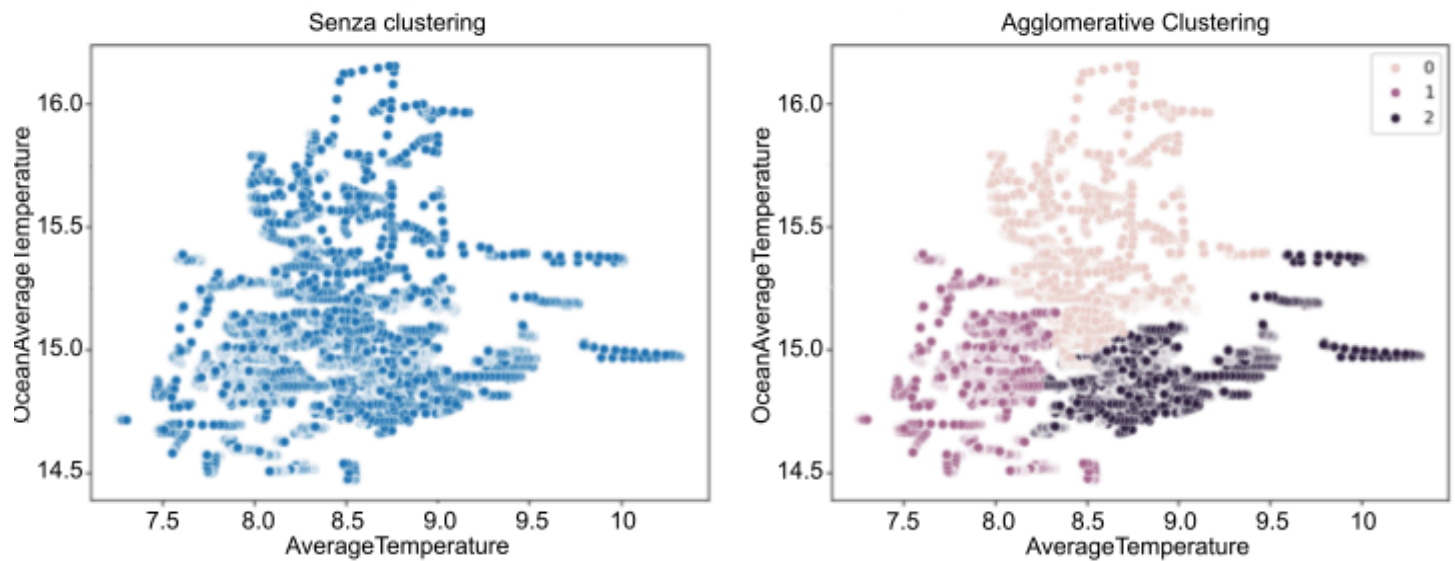


Senza clustering



Agglomerative Clustering

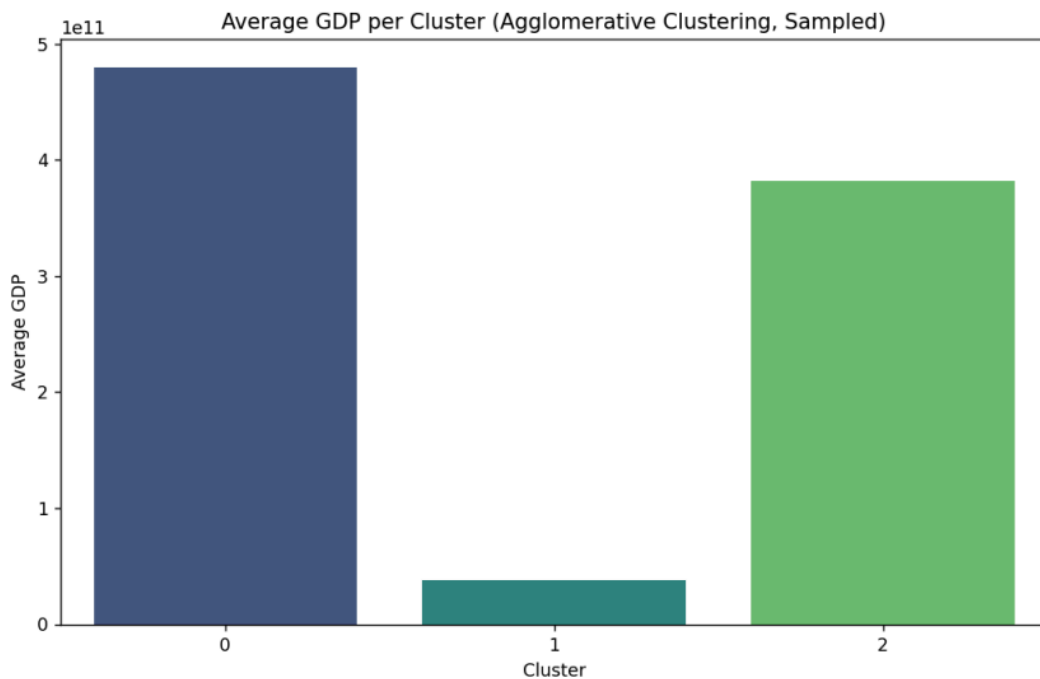




L'Agglomerative Clustering ha prodotto cluster con una chiara distinzione tra regioni con un forte incremento delle temperature medie terrestri (Cluster 0), regioni con un aumento moderato delle temperature (Cluster 1) e regioni con un incremento significativo delle temperature minime terrestri (Cluster 2).

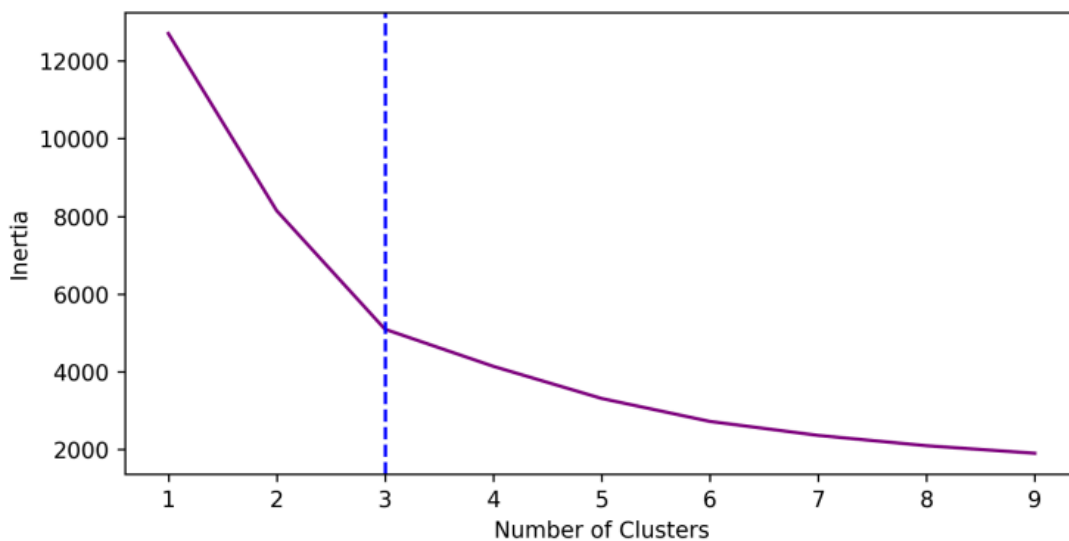
Tuttavia, poiché i principi di classificazione non erano ben definiti, abbiamo considerato altri fattori nell'analisi, come il PIL o GDP (Prodotto Interno Lordo). Integrandolo nell'analisi, siamo giunti a nuove conclusioni:

- Le regioni con un PIL più alto tendono a mostrare un aumento più evidente delle temperature medie, probabilmente a causa di un maggiore sviluppo industriale e urbanizzazione, che contribuiscono alle emissioni di gas serra.
- Le regioni con un PIL più basso, sebbene possano mostrare un aumento meno marcato delle temperature, potrebbero avere dati meno accurati o sottostimati a causa della mancanza di risorse per il monitoraggio climatico e ambientale."



KMEANS CLUSTERING

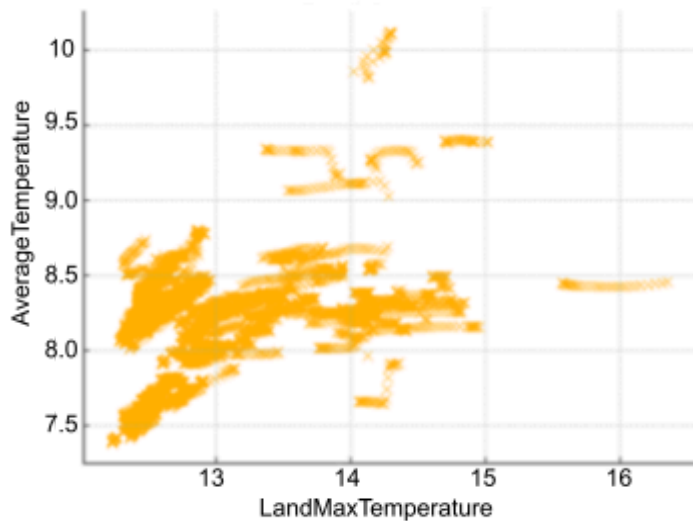
Abbiamo impiegato il metodo dell' "elbow" (gomito) per stabilire il numero ottimale di cluster nell'analisi di clustering k-means. Questo metodo ci ha consentito di individuare il punto in cui aggiungere ulteriori cluster non migliora significativamente la qualità del clustering. Esaminando la somma delle distanze quadrate all'interno dei cluster per vari numeri di cluster, abbiamo determinato che il numero ottimale di cluster è 3.



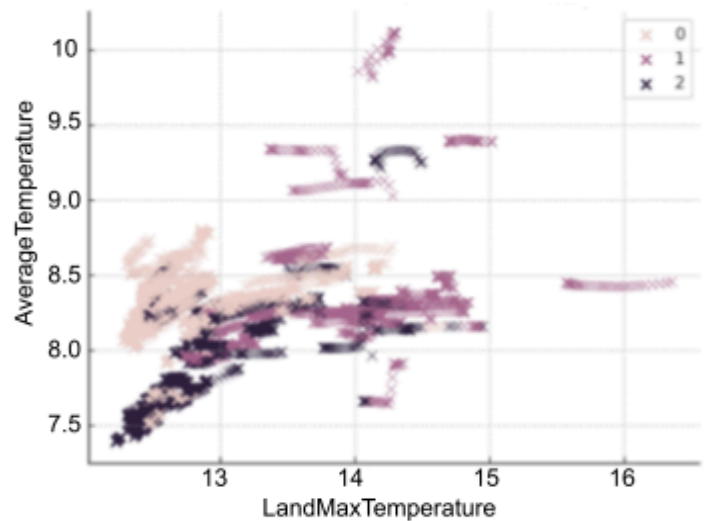
I risultati hanno mostrato che le regioni sono state distribuite in modo simile tra i cluster sia nel clustering agglomerativo che in quello KMeans.

Tuttavia, il coefficiente di silhouette per il clustering KMeans è stato leggermente superiore, raggiungendo un valore di 0.400. Questo indica che il metodo KMeans offre una separazione dei cluster leggermente migliore rispetto al metodo agglomerativo per questo particolare dataset.

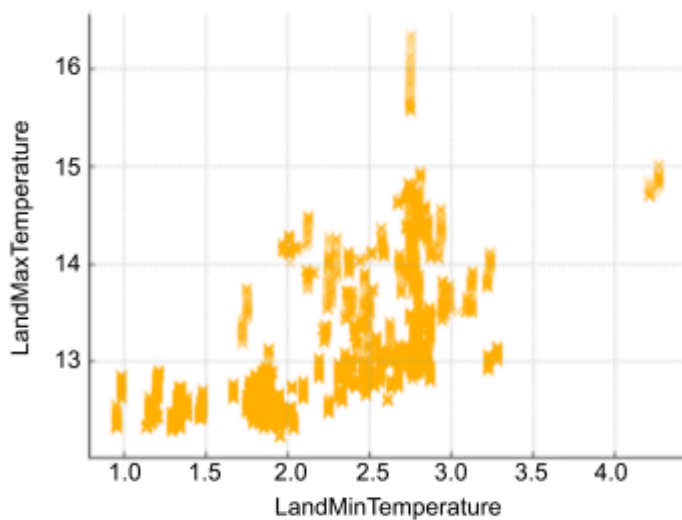
Senza clustering



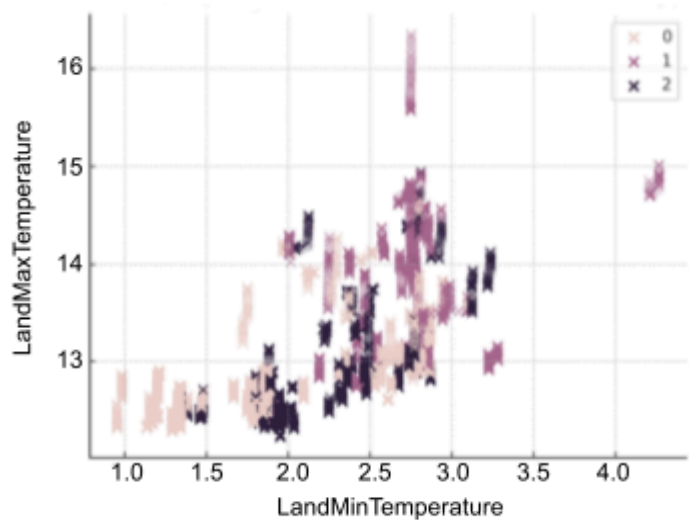
KMeans Clustering



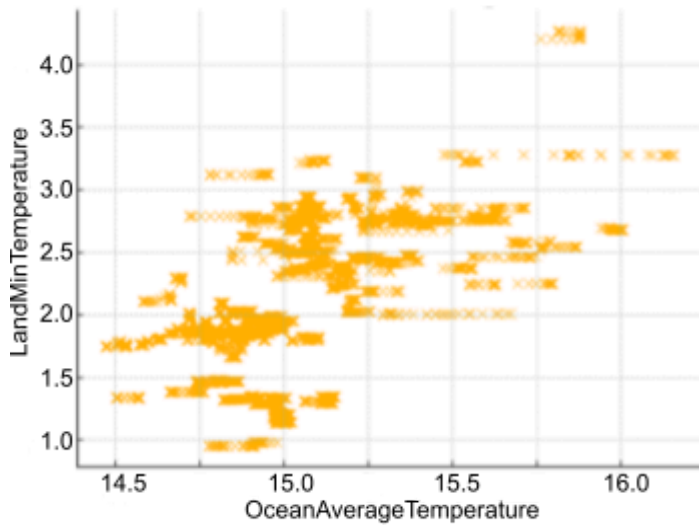
Senza clustering



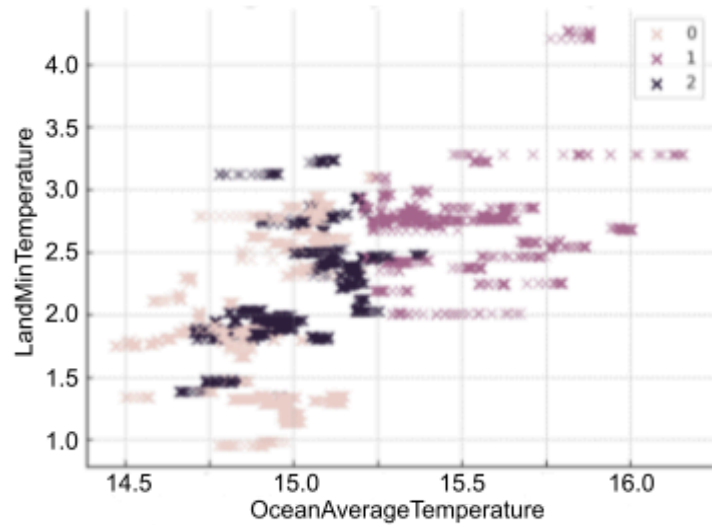
KMeans Clustering



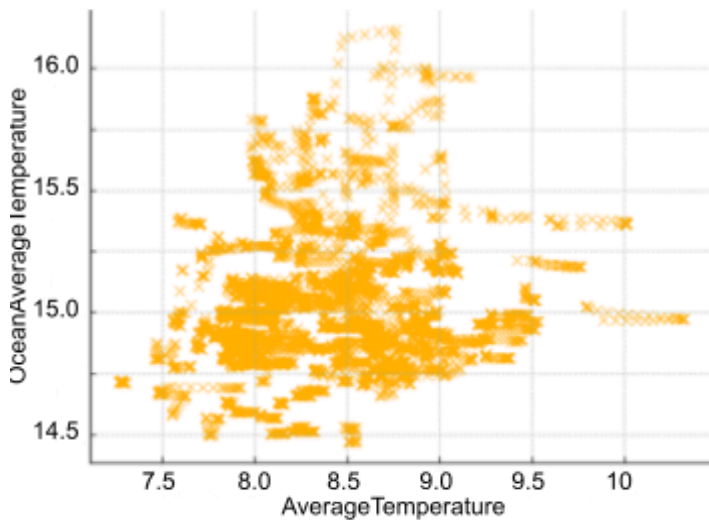
Senza clustering



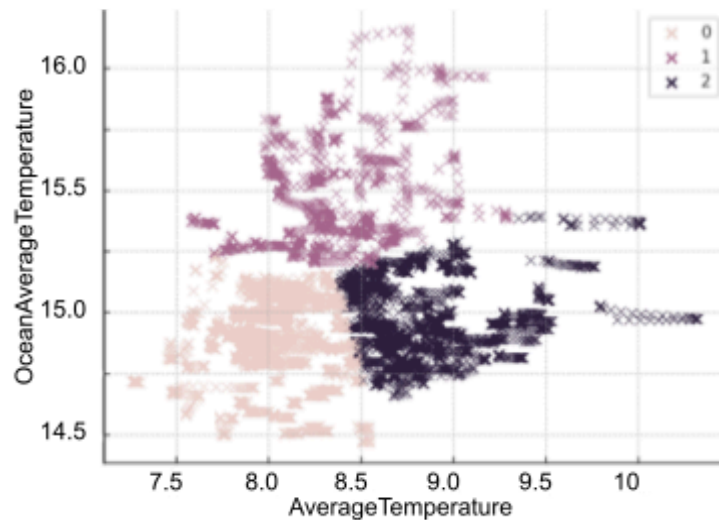
KMeans Clustering



Senza clustering



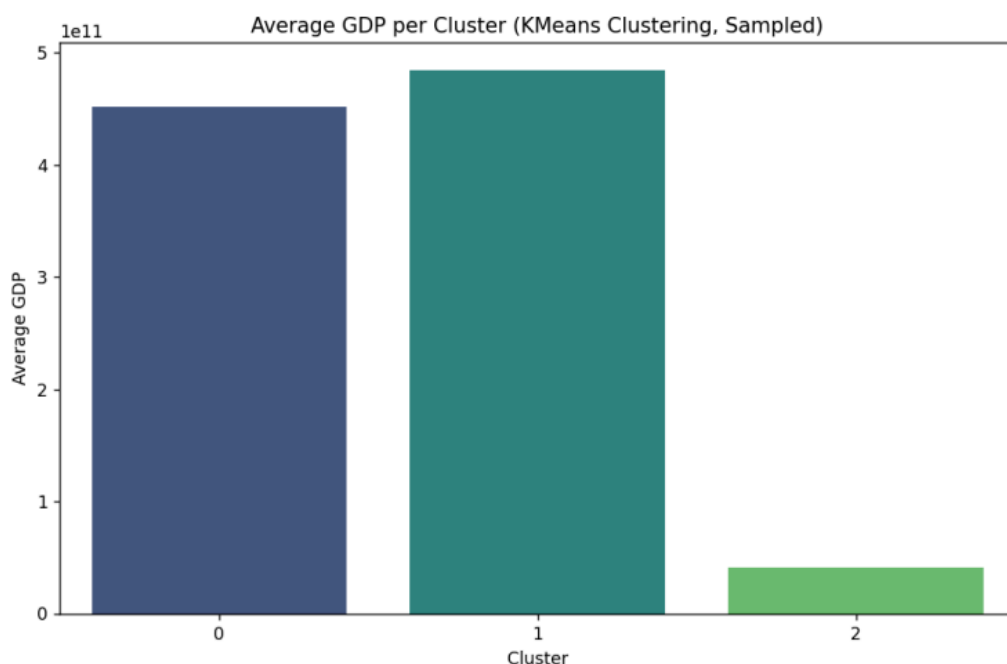
KMeans Clustering



Anche in questo caso, i cluster mostrano una distinzione tra regioni con un forte incremento delle temperature medie terrestri (Cluster 1) e regioni con un incremento significativo delle temperature minime terrestri (Cluster 2):

- **il Cluster 0** include regioni economicamente diverse, combinando paesi occidentali e mediorientali, suggerendo un mix di influenze culturali e ambientali
- **il Cluster 1** è caratterizzato da paesi in via di sviluppo e sviluppati, principalmente dell'Asia e dell'Europa orientale
- **il Cluster 2** è prevalentemente composto da regioni africane, evidenziando sfide economiche e potenziali difficoltà nella gestione e adattamento ai cambiamenti climatici.

Anche in questo caso, poiché i principi di classificazione non erano ben definiti sulla base delle sole temperature, abbiamo considerato altri fattori nell'analisi, in particolare il PIL/GPD



CONCLUSIONI

L'analisi ha rivelato significative disparità economiche tra i cluster, suggerendo che il PIL può influenzare i risultati climatici osservati. Le regioni con un PIL più alto tendono a essere raggruppate insieme, indicando che i fattori economici possono giocare un ruolo importante nel grado di riscaldamento e nelle risposte ai cambiamenti climatici.

Sia nei cluster identificati dal metodo K-means che in quelli dell'agglomerative clustering:

- Le regioni con un PIL elevato mostrano un maggiore incremento delle temperature, suggerendo che le risorse economiche possono influenzare l'industrializzazione, l'urbanizzazione e le emissioni di gas serra, contribuendo al cambiamento climatico.
- Le regioni con un PIL basso potrebbero avere un incremento delle temperature meno documentato o sottostimato a causa di limitazioni nelle infrastrutture per il monitoraggio climatico e nella capacità di adattamento.

METODOLOGIA	N° DI CLUSTER	COEFFICIENTE DI SILHOUETTE	DESCRIZIONE DEL CLUSTER
Clustering Agglomerative	3	0.388	Cluster ragionevolmente separati ma potrebbe esserci qualche sovrapposizione
Kmeans Clustering	3	0.400	Cluster

			ragionevolmente separati ma potrebbe esserci qualche sovrapposizione
--	--	--	--

In base ai risultati dell'analisi, tenendo conto sia dell'efficienza che della qualità della separazione dei cluster, il **KMeans Clustering** è preferibile per questo specifico dataset. Questo metodo ha mostrato un coefficiente di silhouette leggermente superiore, indicando una migliore separazione dei cluster, ed è anche più efficiente dal punto di vista computazionale, rendendolo più adatto per analisi future su dataset di dimensioni maggiori.

Profili individuati:

Cluster	Descrizione	PIL (medio)	Prevalenza di cambiamenti climatici	Proposte di interventi
0	Include nazioni economicamente diverse con uno sviluppo da moderato a elevato, unendo paesi occidentali e mediorientali. Tra cui: Germania, Regno Unito, Francia, Italia, Canada, Paesi Bassi, Arabia Saudita, Emirati Arabi Uniti	\$42.21 miliardi	Significativo aumento delle temperature medie e minime terrestri. Rilevante incremento delle temperature oceaniche.	Potenziamento delle infrastrutture per la mitigazione climatica. Campagne di sensibilizzazione sulla riduzione delle emissioni di gas serra. Supporto alle tecnologie rinnovabili e alle pratiche sostenibili.
1	Include un ampio mix di paesi in via di sviluppo e sviluppati, prevalentemente dell'Asia e dell'Europa orientale. Tra cui: Cina India, Russia, Brasile, Sud Africa, Turchia, Polonia, Indonesia, Malesia, Ucraina	\$48.25 miliardi	Moderato aumento delle temperature medie terrestri. Variazioni meno pronunciate nelle temperature minime e oceaniche.	Promozione di politiche di adattamento climatico e resilienza. Incentivi per l'adozione di energie rinnovabili e tecnologie a bassa emissione di carbonio. Educazione e

				sensibilizzazione pubblica sui cambiamenti climatici
2	Prevalentemente composto da paesi africani.	\$4.84 miliardi	Incremento delle temperature minime terrestri e variazioni moderate nelle temperature medie.	Sviluppo di politiche nazionali di adattamento climatico. Rafforzamento delle capacità di monitoraggio e risposta ai disastri climatici. Programmi di prevenzione e promozione della sostenibilità ambientale.

Ogni cluster presenta diverse sfide che richiedono approcci mirati per affrontare gli effetti del cambiamento climatico. Le proposte di intervento suggerite sono progettate per rispondere alle specifiche esigenze di ciascun cluster, con l'obiettivo di mitigare gli impatti climatici e promuovere la sostenibilità attraverso strategie basate sui dati e sulle condizioni socio-economiche locali. Implementare queste politiche può portare a un significativo miglioramento della resilienza ambientale e ridurre le disparità nella capacità di adattamento tra le diverse regioni globali.

APPRENDIMENTO SUPERVISIONATO

È stato utilizzato l'apprendimento supervisionato con l'obiettivo di prevedere i futuri impatti dei cambiamenti climatici, come l'aumento delle temperature medie e minime terrestri e degli oceani in Italia, utilizzando dati storici. È stata condotta un'analisi comparativa per identificare il modello predittivo più accurato e affidabile, che può essere utilizzato per ulteriori previsioni in altre nazioni.

Preparazione dei dati:

Abbiamo iniziato filtrando il dataset originale per includere solo i dati relativi all'Italia. Successivamente, abbiamo selezionato le colonne di interesse che includono l'anno (Year) e le variabili climatiche rilevanti: temperatura media terrestre, temperatura massima terrestre, temperatura minima terrestre, e temperatura media degli oceani. Infine, abbiamo definito un intervallo di anni futuri (2020-2030) per effettuare le previsioni, consentendo di fornire una visione prospettica su come le condizioni climatiche potrebbero evolvere nel prossimo decennio.

Modelli utilizzati

- **Regressione lineare:** Il modello di regressione lineare presuppone una relazione lineare tra la variabile indipendente (Year) e le variabili dipendenti (temperatura media terrestre, temperatura massima terrestre, temperatura minima terrestre e temperatura media degli oceani). Per valutare le prestazioni del modello, è stata utilizzata la tecnica della cross-validation con 5 fold.

La cross-validation suddivide il dataset in cinque parti (fold), utilizzando quattro di queste per l'addestramento del modello e una per la validazione. Questo processo viene ripetuto cinque volte, cambiando ogni volta la parte utilizzata per la validazione, garantendo che ogni dato venga utilizzato sia per l'addestramento che per il test. Questo modello ha fornito una base di riferimento per le previsioni. Anche se semplice, è utile per capire le tendenze lineari nei dati storici relativi alle variabili climatiche.

- **Regressione polinomiale (grado 2):** Il modello di regressione polinomiale di grado 2 estende la regressione lineare includendo termini polinomiali, permettendo così di catturare relazioni non lineari nei dati.

Anche in questo caso, abbiamo utilizzato la cross-validation con 5 fold e l'RMSE come metrica di valutazione. Questo approccio ci ha permesso di confrontare direttamente la performance tra i due modelli di regressione e valutare l'adeguatezza del modello polinomiale nel catturare le dinamiche più complesse dei dati.

Ha migliorato le previsioni rispetto alla regressione lineare catturando relazioni non lineari ma, può risultare più sensibile agli outlier e alle fluttuazioni nei dati.

- **Random forest:** è particolarmente utile per catturare relazioni complesse e non lineari nei dati grazie alla sua struttura basata su una serie di alberi decisionali. Per ottimizzare le prestazioni del modello, abbiamo utilizzato una ricerca randomizzata dei parametri (RandomizedSearchCV). Questa tecnica esplora una vasta gamma di combinazioni di iperparametri per trovare la configurazione ottimale.

Gli iperparametri ottimizzati includono il numero di alberi (`n_estimators`), la profondità massima degli alberi (`max_depth`), il numero minimo di campioni per dividere un nodo (`min_samples_split`) e il numero minimo di campioni per un nodo foglia (`min_samples_leaf`).

Anche per Random Forest, abbiamo utilizzato la cross-validation con 5 fold e l'RMSE come metrica di valutazione. Questo modello ha generalmente fornito le previsioni più accurate, grazie alla sua capacità di modellare relazioni complesse e non lineari. Tuttavia, è anche il più complesso e computazionalmente intensivo.

Il codice utilizza l'errore quadratico medio (RMSE) come metrica di valutazione delle prestazioni poiché è particolarmente utile in situazioni in cui gli errori grandi sono più penalizzanti rispetto a quelli piccoli.

L'elevazione al quadrato degli errori maggiori rende questa metrica più sensibile a errori significativi, il che è cruciale in contesti come la previsione degli impatti climatici, dove un grande

errore nelle stime può avere importanti implicazioni per la pianificazione e la gestione delle risorse ambientali.

ANALISI DEI RISULTATI

Questi sono i risultati del confronto dei modelli

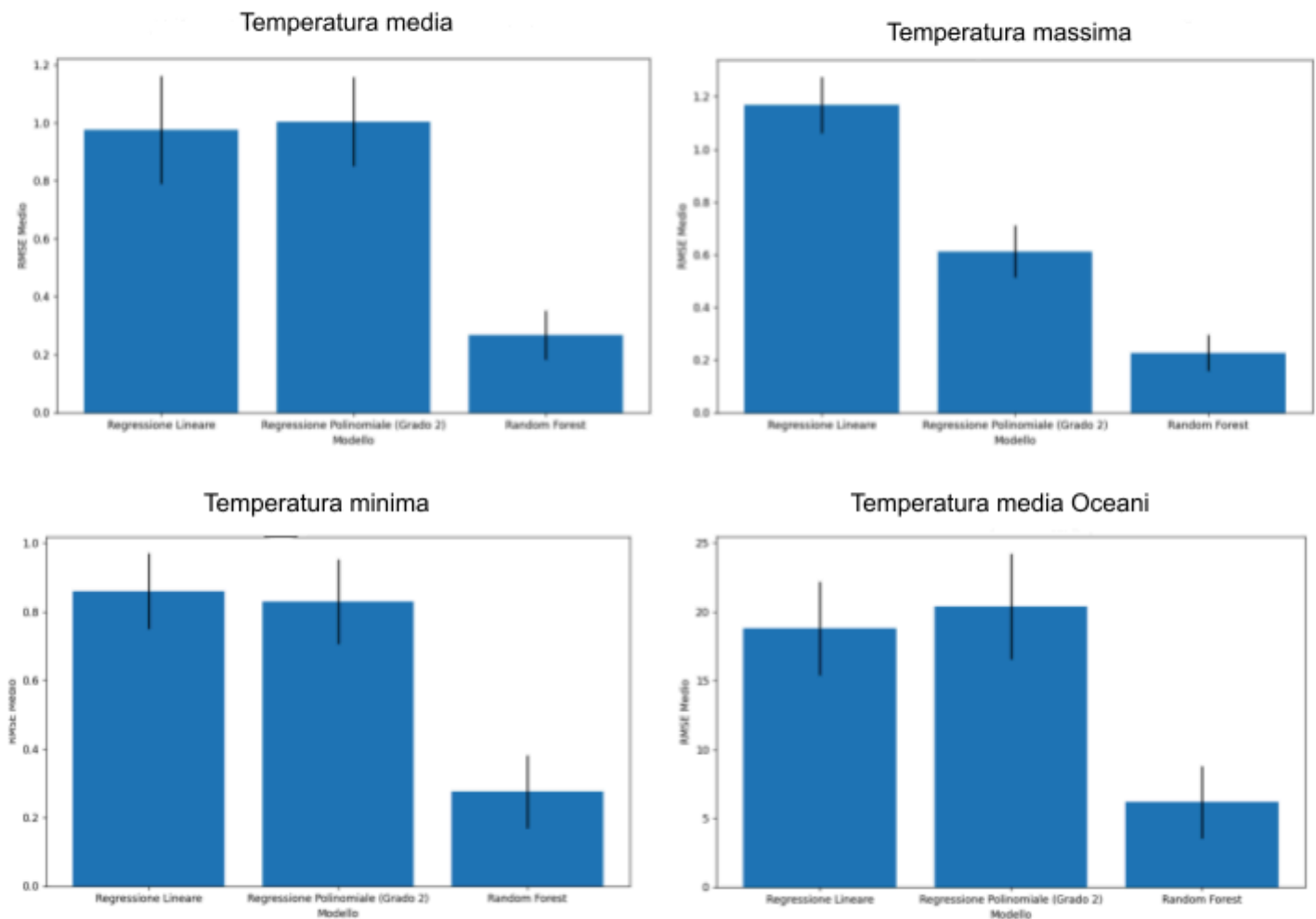
Indice	Modello	Variabile Dipendente	RMSE medio	Deviazione Std RMSE	Migliori Parametri
0	Regressione Lineare	Temperatura Media	0.976	0.186	None
1	Regressione Polinomiale (2)	Temperatura Media	1.004	0.153	None
2	Random Forest	Temperatura Media	0.267	0.085	{'n_estimators': 100, 'min_samples_split': 5}
3	Regressione Lineare	Temperatura Massima	1.169	0.106	None
4	Regressione Polinomiale (2)	Temperatura Massima	0.612	0.099	None
5	Random Forest	Temperatura Massima	0.227	0.069	{'n_estimators': 200, 'min_samples_split': 2}
6	Regressione Lineare	Temperatura Minima	0.859	0.109	None
7	Regressione Polinomiale (2)	Temperatura Minima	0.829	0.123	None
8	Random Forest	Temperatura Minima	0.275	0.106	{'n_estimators': 200, 'min_samples_split': 5}
9	Regressione Lineare	Temp. Media Oceani	18.802	3.400	None

10	Regressione Polinomiale (2)	Temp. Media Oceani	20.392	3.840	None
11	Random Forest	Temp. Media Oceani	6.186	2.618	{'n_estimators': 200, 'min_samples_split': 2}

In genere, i modelli Random Forest superano i modelli di regressione lineare e polinomiale in termini di RMSE inferiore, indicando prestazioni predittive migliori.

La regressione polinomiale potrebbe talvolta mostrare una maggiore variabilità (deviazione standard) in RMSE a causa della sua sensibilità al grado del polinomio e al potenziale adattamento eccessivo su set di dati di piccole dimensioni.

I grafici a barre per ciascun cambiamento climatico mostrano l'RMSE medio per i tre modelli, fornendo un confronto visivo delle loro prestazioni. Le barre inferiori indicano prestazioni migliori (errore inferiore).



APPROCCIO ENSEMBLE

Viene utilizzato un approccio ensemble per combinare le previsioni dei tre modelli. I pesi per ciascun modello sono inversamente proporzionali ai rispettivi RMSE, fornendo una previsione equilibrata che sfrutta i punti di forza di ciascun modello.

I risultati del Confronto dei **Modelli Ensemble** sono i seguenti:

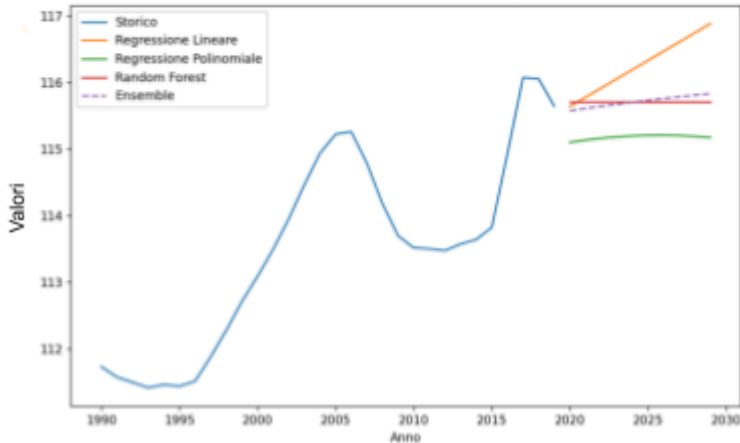
Indice	Variabile	RMSE Medio Ensemble	Deviazione Std RMSE Ensemble
0	Temperatura Media	1.061	0.215
1	Temperatura Massima	1.055	0.213
2	Temperatura Minima	0.762	0.179
3	Temp. Media Oceani	17.498	3.013

Questo modello ha mostrato prestazioni miste.

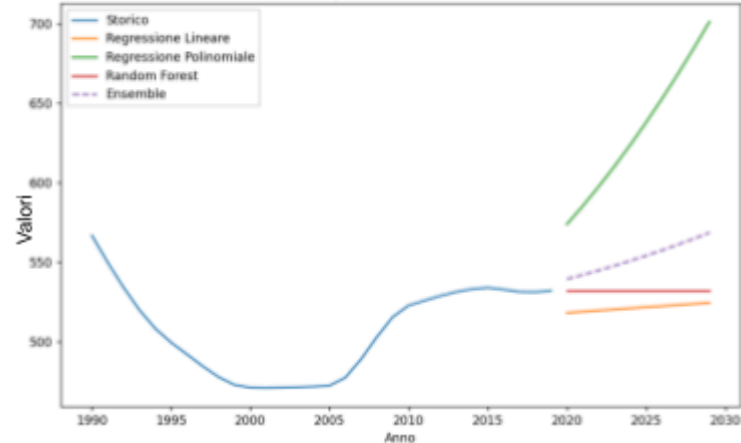
Sebbene abbia fornito una buona accuratezza per alcune variabili climatiche, come la temperatura massima e la temperatura minima, non è riuscito a superare significativamente le prestazioni del Random Forest per altre variabili.

Ad esempio, per la temperatura massima terrestre, il modello ha avuto un RMSE più elevato rispetto al Random Forest, indicando una minore precisione.

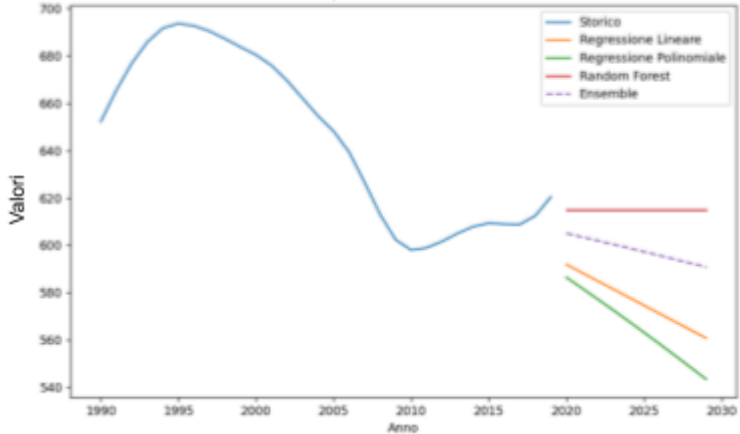
Temperatura Minima



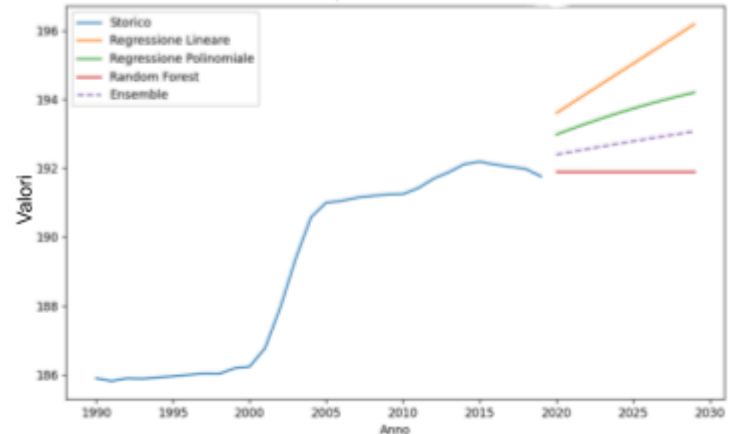
Temperatura Media



Temp. Media Oceani



Temperatura Massima



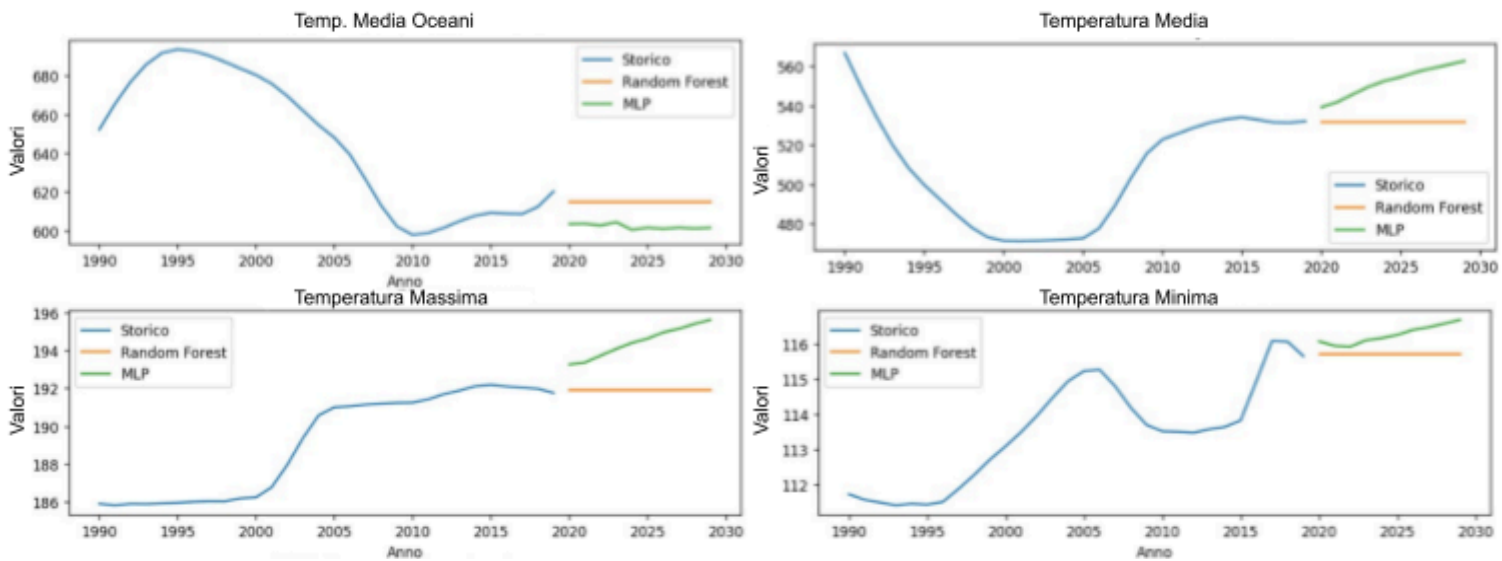
RANDOM FOREST VS MLP (MULTI LAYER PERCEPTRON)

Dall'analisi comparativa tra i modelli Random Forest e MLP (Multi-Layer Perceptron) per la previsione delle variabili climatiche in Italia, emergono alcuni punti chiave. In primo luogo, per tutte le variabili considerate (temperatura media, temperatura massima terrestre, temperatura minima terrestre, temperatura media degli oceani), il modello Random Forest ha mostrato un errore quadratico medio (RMSE) inferiore rispetto al modello MLP, indicando una maggiore precisione nelle previsioni. Analogamente, l'errore assoluto medio (MAE) è stato costantemente inferiore per il modello Random Forest rispetto al MLP.

Per quanto riguarda le performance del modello, il coefficiente di determinazione (R^2) ha evidenziato valori negativi per entrambi i modelli in alcuni casi, suggerendo che non riescono a spiegare adeguatamente la variabilità dei dati. Questo è particolarmente critico per il modello MLP, che ha mostrato prestazioni peggiori rispetto al Random Forest. Tuttavia, il modello Random

Forest ha ottenuto valori R^2 leggermente migliori rispetto al MLP, sebbene ancora negativi, sottolineando la difficoltà intrinseca nella previsione di questi dati complessi.

Variabile	Random Forest RMSE	MLP RMSE	Random Forest MAE	MLP MAE	Random Forest R^2	MLP R^2	Random Forest MAPE	MLP MAPE
Temp. Media Oceani	10.131	8.693	8.973	7.189	-1.557	-0.882	0.014	0.011
Temperatura massima	1.200	1.105	0.908	0.934	-1.150	-1.132	0.004	0.004
Temperatura media	3.040	2.696	2.241	2.611	-0.694	-1.436	0.001	0.011
Temperatura minima	1.678	2.051	1.423	1.829	-1.460	-2.555	0.012	0.016



OTTIMIZZAZIONE DEL RANDOM FOREST

Per migliorare le performance del modello, soprattutto in termini di R^2 , è stata implementata una nuova strategia di modellazione utilizzando Random Forest con un approccio più sofisticato.

In primo luogo, è stata effettuata una selezione delle feature, includendo tutte le metriche disponibili nel dataset e aggiungendo l'anno come variabile di input.

Questo passaggio è cruciale perché l'inclusione di variabili aggiuntive può fornire al modello informazioni utili per migliorare la precisione delle previsioni.

Per garantire una valutazione robusta del modello, è stata utilizzata la tecnica della **cross-validation** con K-Fold. I risultati della cross-validation hanno fornito valori medi e deviazioni standard del RMSE, offrendo un'ulteriore misura della precisione del modello.

Successivamente, la scala delle variabili è stata normalizzata utilizzando lo **StandardScaler**, migliorando così la capacità del modello di apprendere dalle diverse caratteristiche con scale variabili.

Il modello Random Forest è stato quindi addestrato e testato su un set di dati separati, con la valutazione delle sue prestazioni tramite metriche come R^2 , MSE (Mean Squared Error), RMSE e MAE.

Per rendere il modello più robusto, sono stati utilizzati gli ultimi dati disponibili per l'Italia, **aggiornando continuamente le feature con nuove previsioni annuali**.

Questo processo iterativo ha permesso di generare stime per le variabili climatiche future, tenendo conto anche di un leggero rumore aggiunto ai dati, simulando la variabilità e l'incertezza tipiche dei dati reali.

Le previsioni ottenute sono state visualizzate graficamente e hanno mostrato una tendenza coerente con i dati storici, fornendo una chiara indicazione dei possibili scenari futuri. I risultati finali sono stati poi salvati in un file CSV per ulteriori analisi e confronti.

Questi sono i risultati ottenuti:

Valori di valutazione per Temp. Media Oceani:

R²: 0.994026265100252

MSE: 6.31029462772926

RMSE: 2.512029551433049

MAE: 2.55182683338883

Cross-Validation RMSE Mean: 4.391850077435272

Cross-Validation RMSE Std: 0.910561028535973

Valori di valutazione per Temperatura massima:

R²: 0.966708931880889

MSE: 0.1655916825127384

RMSE: 0.3249487382843236

MAE: 0.4522429999999999

Cross-Validation RMSE Mean: 0.28320515351668085

Cross-Validation RMSE Std: 0.1218093835110354

Valori di valutazione per Temperatura media:

R²: 0.987295887032145

MSE: 0.08467681558357053

RMSE: 0.29099281019222883

MAE: 0.2009695200000426

Cross-Validation RMSE Mean: 0.45512259143537814

Cross-Validation RMSE Std: 0.25227797463123225

Valori di valutazione per Temperatura minima:

R²: 0.9464360217444229

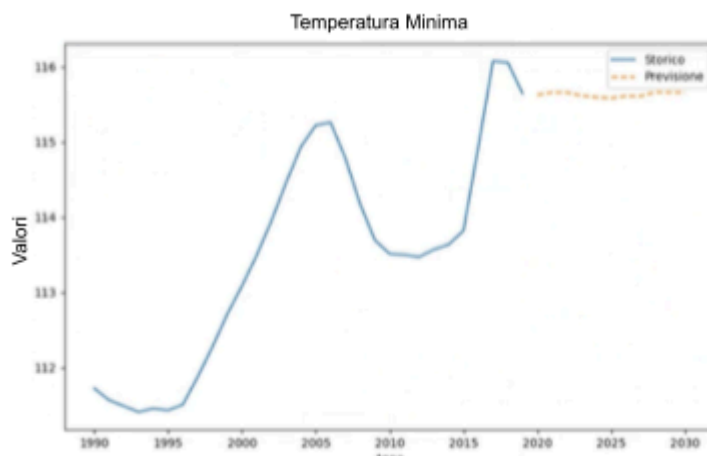
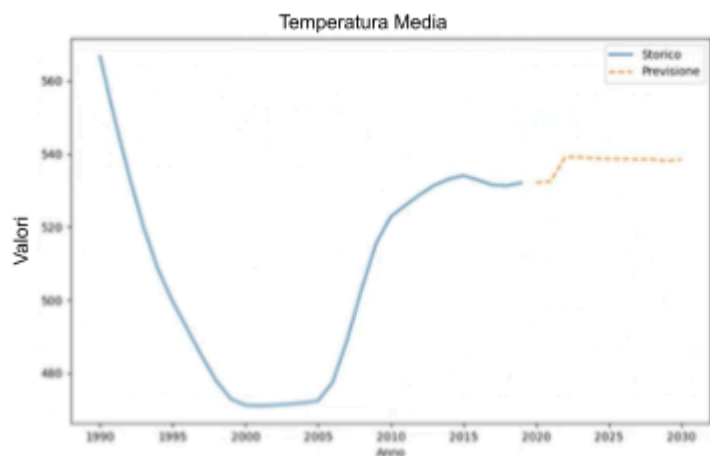
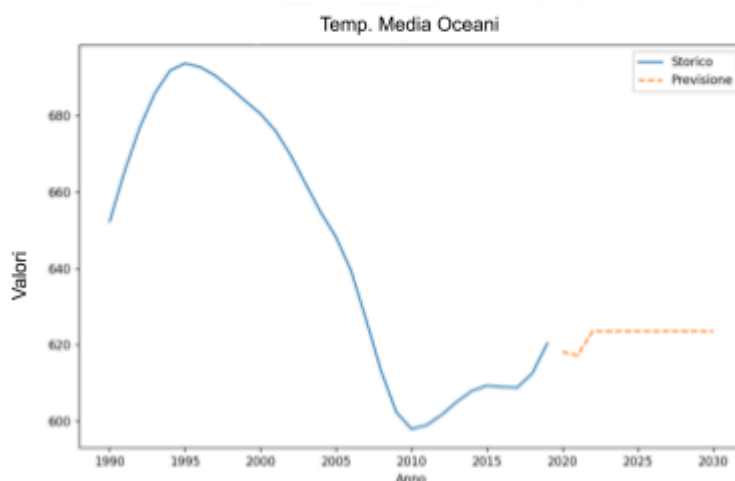
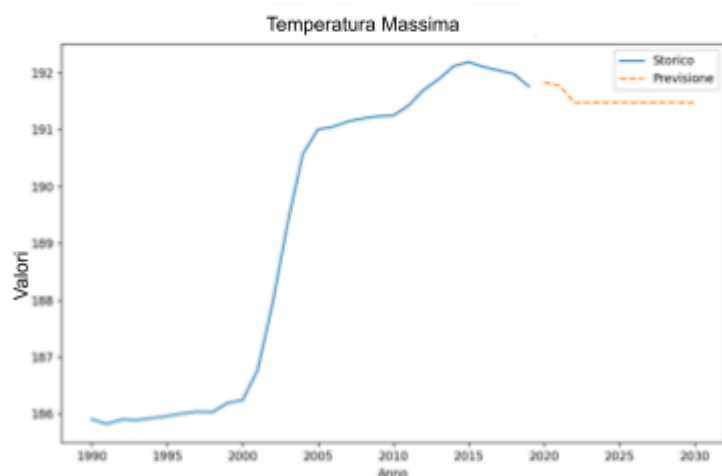
MSE: 0.09976514786432082

RMSE: 0.3158419645975863

MAE: 0.2604410500000498

Cross-Validation RMSE Mean: 0.29890463772206995

Cross-Validation RMSE Std: 0.1106700947911922



CONCLUSIONI

La previsione delle variabili climatiche in Italia ha rivelato diverse tendenze importanti:

1. **Temperatura Media:** Le previsioni indicano un possibile lieve aumento delle temperature medie nei prossimi anni. Questo trend potrebbe essere influenzato da fattori globali come l'aumento dei gas serra e il cambiamento delle correnti oceaniche.
2. **Temperatura Massima Terrestre:** Le proiezioni mostrano un aumento significativo delle temperature massime terrestri, suggerendo che l'Italia potrebbe affrontare estati sempre più calde. Questo incremento può avere implicazioni gravi per l'agricoltura, la salute pubblica e l'energia.
3. **Temperatura Minima Terrestre:** Anche le temperature minime terrestri sono previste in aumento, sebbene in misura minore rispetto alle massime. Questo potrebbe portare a inverni più miti, con conseguenze per gli ecosistemi locali e la biodiversità.

4. **Temperatura Media degli Oceani:** Le previsioni indicano un aumento della temperatura media degli oceani, che potrebbe influenzare i pattern meteorologici e la biodiversità marina, nonché il livello del mare lungo le coste italiane.

Proposte di Piani di Adattamento e Mitigazione:

Sulla base delle raccomandazioni dell'Organizzazione Mondiale della Meteorologia e altre linee guida internazionali, abbiamo formulato i seguenti piani d'intervento:

- **Gestione delle Risorse Idriche:** Implementare strategie di gestione dell'acqua per affrontare la crescente domanda durante i periodi di siccità previsti. Questo include l'adozione di tecnologie per il risparmio idrico, il miglioramento delle infrastrutture di stoccaggio e distribuzione dell'acqua, e la promozione di pratiche agricole sostenibili.
- **Pianificazione Urbana:** Le città devono adattarsi al riscaldamento globale mediante la progettazione di infrastrutture resilienti. Questo include la creazione di spazi verdi urbani per ridurre l'isola di calore urbana e la promozione di edifici ad alta efficienza energetica.
- **Agricoltura e Sicurezza Alimentare:** Investire in tecnologie agricole avanzate e resilienti ai cambiamenti climatici, come colture resistenti alla siccità e metodi di irrigazione efficienti, per mantenere la produttività agricola e la sicurezza alimentare.
- **Ricerca e Monitoraggio:** Creare un database nazionale per monitorare i cambiamenti climatici e i loro impatti. Questo include la raccolta di dati su temperature, precipitazioni, e altre variabili climatiche per facilitare l'aggiornamento continuo delle strategie di adattamento.

Questi piani sono fondamentali per migliorare la resilienza dell'Italia ai cambiamenti climatici, proteggendo al contempo la qualità della vita e riducendo le disparità socio-economiche.

ONTOLOGIA

Abbiamo utilizzato i seguenti strumenti:

- **pandas:** per caricare e manipolare il dataset
- **rdflib:** per caricare e interrogare l'ontologia RDF
- **protégé:** per visualizzare la struttura dell'ontologia

Abbiamo sviluppato un'ontologia per organizzare i nostri dati e risultati sui cambiamenti climatici, comprendendo variabili come le temperature medie, massime e minime, nonché le temperature medie degli oceani. L'obiettivo è facilitare la condivisione, il riutilizzo e l'interoperabilità dei dati tra diversi sistemi e applicazioni. Questo approccio consente anche l'utilizzo di inferenze e ragionamenti automatici, migliorando l'analisi e la comprensione dei dati.

L'integrazione della nostra ontologia con altre ontologie ambientali standard, come la "Environment Ontology (ENVO)", ci permette di utilizzare definizioni e relazioni standardizzate esistenti. Ciò assicura che i nostri dati siano compatibili con altre fonti di dati e strumenti che

adottano la stessa ontologia, promuovendo una maggiore coerenza e interoperabilità nella gestione dei dati climatici.

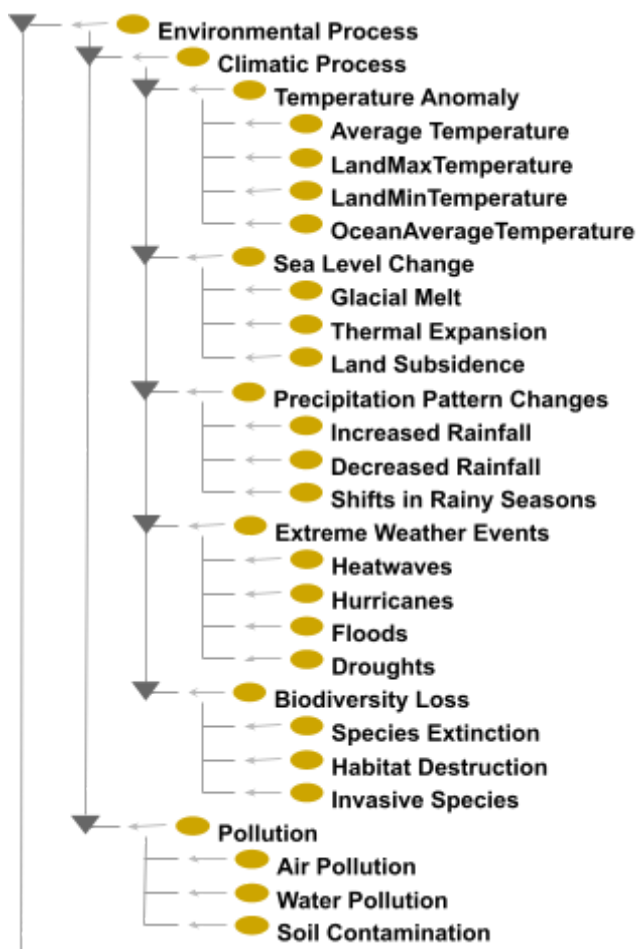
ENVIRONMENT ONTOLOGY (ENVO)

“L'ontologia ambientale, conosciuta come "Environment Ontology (ENVO)", è stata sviluppata come uno standard per descrivere le caratteristiche e le condizioni ambientali. Il suo scopo è fornire alla comunità scientifica descrizioni coerenti, riutilizzabili e sostenibili dei termini legati all'ambiente, attraverso sforzi collaborativi con ricercatori e istituzioni accademiche.”

Per iniziare, abbiamo esplorato la struttura dell'ontologia utilizzando Protégé, un noto strumento di sviluppo per ontologie.

Climatic Process, una sottoclasse di Environmental Process, include vari sottoprocessi climatici rilevanti per la nostra analisi.

Ogni categoria principale è suddivisa in sottoclassi che descrivono vari fenomeni climatici specifici:



Temperature Anomaly è suddivisa in diverse forme, ciascuna con caratteristiche distintive relative alle variazioni di temperatura rispetto ai valori storici medi.

Sea Level Change include varie forme di cambiamenti del livello del mare, causati da fenomeni come lo scioglimento dei ghiacci polari e la dilatazione termica degli oceani.

Precipitation Pattern Changes comprende differenti tipi di variazioni nei modelli di precipitazione, che possono influenzare sia la quantità sia la distribuzione delle precipitazioni in diverse regioni.

Extreme Weather Events copre una gamma di eventi meteorologici estremi, come uragani, ondate di calore e inondazioni, ognuno con manifestazioni uniche e impatti potenzialmente devastanti.

Biodiversity Loss include diversi tipi di perdita di biodiversità, con una gamma di effetti sugli ecosistemi e sulle specie.

METODOLOGIA

Per garantire coerenza e evitare conflitti di denominazione, abbiamo definito due namespace per facilitare l'integrazione con altre ontologie e dataset:

- **CLIMATE**: per la nostra ontologia personalizzata, che organizza i dati relativi ai cambiamenti climatici.
- **ENVO**: per l'Environment Ontology, che fornisce un quadro standardizzato per la descrizione delle variabili ambientali.

Abbiamo mappato le variabili climatiche presenti nel nostro dataset con gli **URI** (Uniform Resource Identifier) corrispondenti dell'Environment Ontology. Ad esempio, la "temperatura media della superficie terrestre" è stata mappata all'URI

`http://purl.envolibrary.org/obo/ENVO_09200012`, la "temperatura media dell'oceano" all'URI
`http://purl.envolibrary.org/obo/ENVO_09200013`, e così via.

Questo approccio consente di utilizzare termini e classificazioni standardizzati, facilitando l'interoperabilità e l'integrazione con altre fonti di dati e strumenti che adottano la stessa ontologia.

Abbiamo creato un **grafo RDF** per rappresentare le informazioni. Il grafo è una struttura dati che contiene nodi (individui) e archi (relazioni) che collegano questi nodi.

Abbiamo quindi creato delle proprietà per collegare le variabili climatiche alle regioni presenti nel dataset. Ad esempio, per rappresentare la temperatura media della superficie terrestre in Italia, abbiamo creato una tripla che collega l'individuo "Italia" con la proprietà "haTemperaturaMedia" e l'oggetto "temperaturaMediaSuperficieTerrestre".

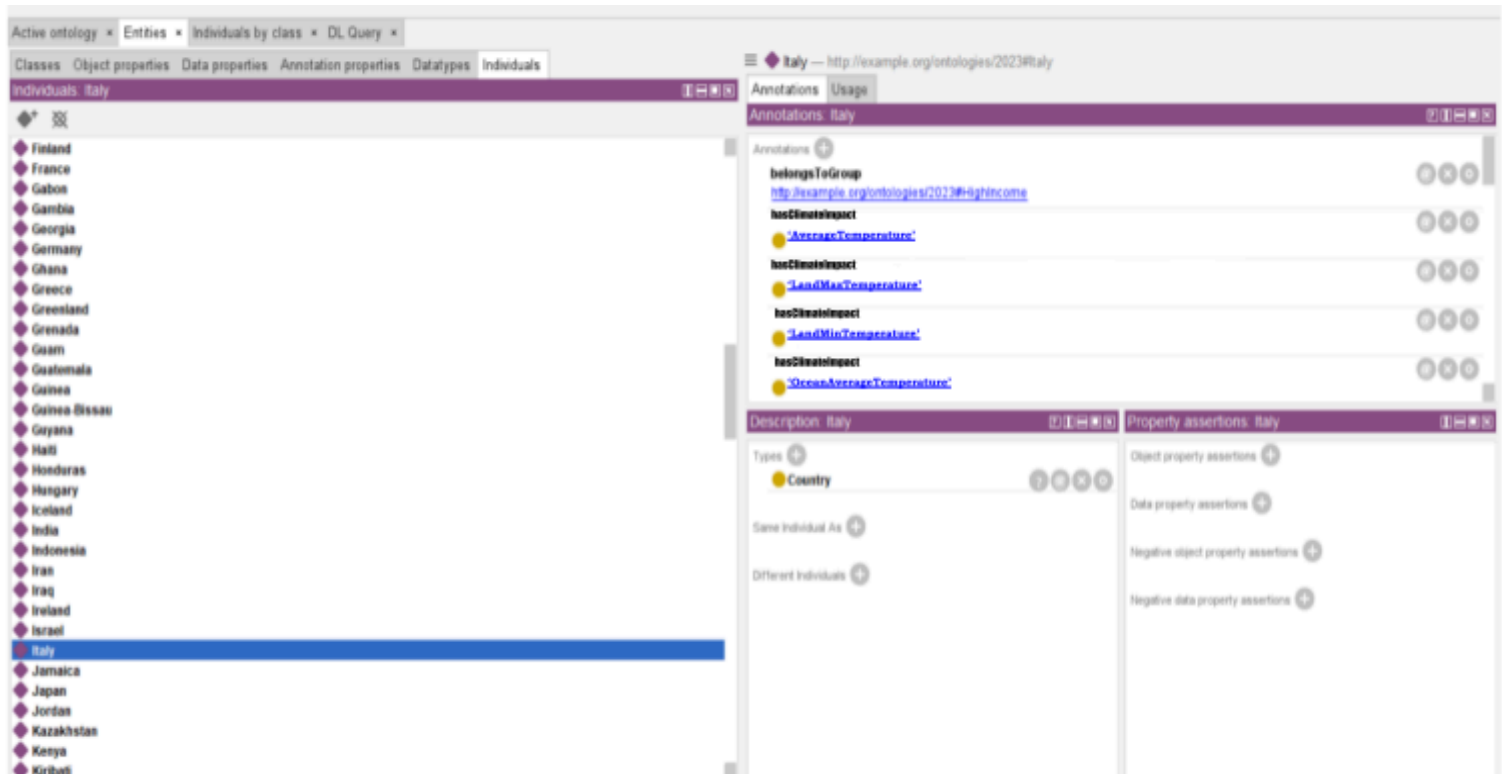
Per ogni riga del dataset, abbiamo generato triple RDF che collegano le regioni (**soggetti**) con vari attributi (**predicati e oggetti**) come l'anno, i dati climatici (es. temperatura media, temperatura minima, temperatura massima) e le statistiche correlate. Ogni tripla RDF rappresenta una singola affermazione riguardante un'entità nel nostro dominio, permettendo una rappresentazione strutturata e interconnessa delle informazioni climatiche.

Il grafo RDF è stato serializzato in un file OWL (**IntegratedOntology.owl**).

La serializzazione è il processo di conversione di una struttura dati in un formato che può essere facilmente memorizzato e recuperato.

Il formato OWL (Web Ontology Language) permette di rappresentare ontologie complesse in modo interoperabile.

Abbiamo verificato la struttura dell'ontologia generata per assicurarci che tutte le classi, le proprietà e gli individui fossero correttamente definiti. Questo passaggio è cruciale per garantire che le informazioni siano rappresentate in modo accurato.



Esempio di utilizzo

Selezionando windows>tabs>SPARQL Query su Protege possiamo effettuare un query SPARQL per interrogare l'ontologia.

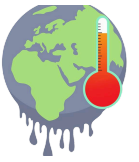
Query di esempio:

```
PREFIX climate: <http://ecoMachineLearning.org/ontologies/2023#>
PREFIX envo: http://purl.envo.library.org/envo/>
SELECT ?region ?averageTemperature
WHERE {
    ?region climate:hasClimateImpact envo:ENVO_00002009 ;
        climate:averageTemperature ?averageTemperature .
    FILTER(?averageTemperature > 15)
}
```

Risultato:un elenco di regioni e le rispettive temperature medie terrestri superiori a 20°C.

oppure

```
PREFIX climate: <http://ecoMachineLearning.org/ontologies/2023#>
SELECT ?region ?temperatureGroup
WHERE {
    ?region climate:belongsToTemperatureGroup ?temperatureGroup .
}
```



Risultato: un elenco di tutte le regioni con i relativi gruppi di temperature.

CONCLUSIONI FINALI

Le analisi globali indicano che i cambiamenti climatici variano significativamente tra le diverse regioni, influenzati da fattori geografici, economici e di sviluppo sostenibile. Paesi come gli Stati Uniti, la Germania e il Brasile mostrano tendenze crescenti di riscaldamento, sottolineando l'importanza di monitorare e intervenire tempestivamente. Inoltre, l'apprendimento non supervisionato ha portato all'identificazione di cluster di regioni con profili climatici simili, utilizzando tecniche di KMeans Clustering. L'integrazione del PIL nell'analisi ha rivelato disparità economiche significative tra i cluster, suggerendo che le risorse economiche giocano un ruolo cruciale nell'adattamento e mitigazione dei cambiamenti climatici.

Le previsioni future delle temperature medie globali, ottenute tramite Random Forest, indicano una tendenza al rialzo delle temperature medie annuali in molte regioni, evidenziando la necessità di pianificare interventi mirati per affrontare l'impatto crescente del riscaldamento globale.

IDEE PER ESTENSIONI FUTURE

- Utilizzare modelli di serie temporali per migliorare le previsioni delle temperature future.
- Studiare l'impatto di eventi significativi (ad esempio, crisi economiche, pandemie) sui trend climatici.

FINE