MentalHealthAnalytics

Analisi dei disturbi mentali attraverso tecniche di data analysis, machine learning e modellazione predittiva.

Autore
Annarita Bruno, 766402, a.bruno113@studenti.uniba.it

< https://github.com/annabr98/consegnaICON2526.git**>**

AA 2025-2026

INDICE

INTRODUZIONE

Negli ultimi anni la salute mentale ha assunto un ruolo sempre più centrale nel dibattito scientifico e nelle politiche sanitarie, dopo un lungo periodo di relativa sottovalutazione. Il presente lavoro analizza alcune delle principali patologie mentali, tra cui la schizofrenia, i disturbi depressivi, i disturbi d'ansia, i disturbi bipolari e i disturbi del comportamento alimentare.

In una fase iniziale sono stati esaminati dati storici relativi alla diffusione di tali patologie, con l'obiettivo di analizzarne l'evoluzione nel tempo e di individuare eventuali tendenze significative. Questa analisi esplorativa ha permesso di ottenere una visione complessiva della situazione attuale e delle dinamiche passate, supportata dall'utilizzo di grafici e strumenti di visualizzazione dei dati.

Successivamente sono state applicate tecniche di apprendimento non supervisionato per individuare gruppi di paesi con caratteristiche simili in termini di incidenza delle patologie mentali analizzate. Questo approccio ha consentito di identificare cluster omogenei di nazioni, facilitando la definizione di strategie di intervento differenziate e mirate.

Parallelamente sono state adottate tecniche di apprendimento supervisionato per la stima dei DALYs (Disability-Adjusted Life Years), un indicatore utilizzato per misurare l'impatto complessivo delle malattie sulla popolazione. I DALYs permettono di integrare sia gli anni di vita persi a causa di mortalità precoce sia gli anni vissuti con disabilità, fornendo una misura sintetica della gravità delle patologie considerate.

Infine, i dati e i risultati ottenuti sono stati organizzati attraverso un modello ontologico, con l'obiettivo di favorire la strutturazione, la condivisione e l'interoperabilità delle informazioni. L'integrazione con la Human Disease Ontology ha permesso di utilizzare definizioni e relazioni standardizzate già esistenti, rendendo i dati compatibili con altre fonti e strumenti che adottano la stessa ontologia e migliorando la validità complessiva dell'analisi.

SOMMARIO

Il progetto realizza un sistema di analisi della salute mentale che integra tecniche di data analysis, apprendimento automatico e rappresentazione della conoscenza all'interno di un'unica struttura. I dati sulla prevalenza dei disturbi mentali e sui DALYs, provenienti da fonti Kaggle e World Bank, vengono prima sottoposti a preprocessing e analisi esplorativa per individuare trend temporali e differenze tra paesi.
Successivamente, un modulo di apprendimento supervisionato consente la previsione dei DALYs per l'Italia, mentre un modulo di apprendimento non supervisionato identifica cluster di nazioni con profili simili, integrando anche il fattore socio-economico del GDP.
I risultati ottenuti vengono infine organizzati in una knowledge base ontologica basata su RDF/OWL e collegata alla Human Disease Ontology, al fine di garantire interoperabilità, riuso dei dati e coerenza semantica. Questa integrazione permette di rappresentare in modo strutturato le relazioni tra disturbi mentali, paesi e indicatori temporali, supportando l'analisi e la condivisione della conoscenza.

ARGOMENTO 1 – APPRENDIMENTO E INCERTEZZA

SOMMARIO

Per la previsione del carico futuro dei disturbi mentali (depressivi, schizofrenici, bipolari, alimentari e d'ansia) in Italia sono state applicate tecniche di apprendimento supervisionato.
Il dataset è stato filtrato considerando esclusivamente i dati relativi all'Italia e selezionando come variabili l'anno e i valori dei DALYs per ciascun disturbo mentale. È stato definito un intervallo temporale futuro (2020–2030) per stimare l'evoluzione dei DALYs nel prossimo decennio.
I modelli utilizzati sono regressione lineare, regressione polinomiale di grado 2 e Random Forest, valutati tramite cross-validation a 5 fold utilizzando l'RMSE come metrica principale.

STRUMENTI UTILIZZATI

Sono stati utilizzati Python e le librerie pandas e scikit-learn per la preparazione dei dati, l'addestramento dei modelli e la valutazione tramite cross-validation.
Per l'ottimizzazione del modello Random Forest è stata impiegata la procedura di Randomized Search, mentre la standardizzazione delle variabili è stata effettuata tramite StandardScaler.

DECISIONI DI PROGETTO

Il dataset è stato filtrato considerando esclusivamente i dati relativi all'Italia, al fine di effettuare previsioni specifiche sul contesto nazionale.
Sono state selezionate come variabili di input l'anno (Year) e i valori dei DALYs per ciascun disturbo mentale. È stato definito un intervallo temporale futuro (2020–2030) per stimare l'evoluzione dei DALYs nel prossimo decennio.
Per la valutazione dei modelli è stata utilizzata la cross-validation a 5 fold e l'RMSE come metrica principale di confronto. Nel caso del modello Random Forest è stata applicata una procedura di ottimizzazione tramite Randomized Search sugli iperparametri principali.

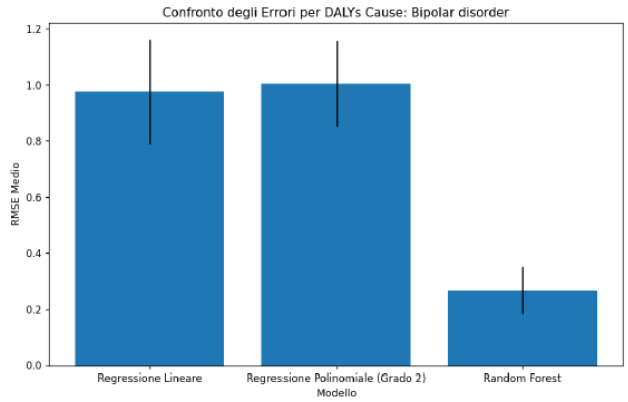
VALUTAZIONE

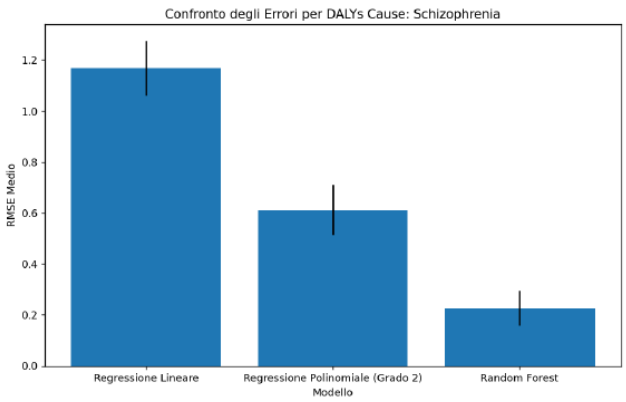
ANALISI DEI RISULTATI

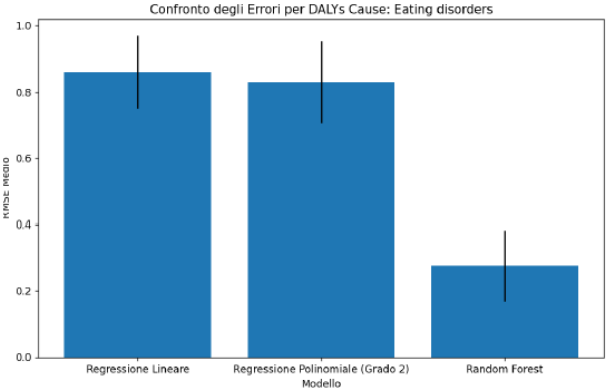
Risultati del Confronto dei Modelli

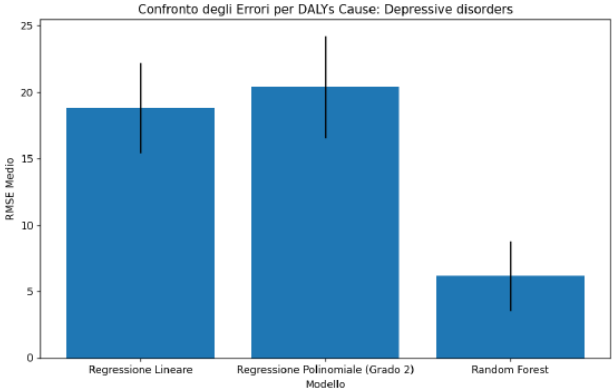
	Modello	Disturbo	RMSE Medio	Deviazione Std RMSE	Migliori Parametri
0	Regressione Lineare	DALYs Cause: Depressive disorders	18.802480	3.400771	None
1	Regressione Polinomiale (Grado 2)	DALYs Cause: Depressive disorders	20.392532	3.840496	None
2	Random Forest	DALYs Cause: Depressive disorders	6.186234	2.618246	{'n_estimators': 100, 'min_samples_split': 5, ...
3	Regressione Lineare	DALYs Cause: Schizophrenia	1.169332	0.106849	None
4	Regressione Polinomiale (Grado 2)	DALYs Cause: Schizophrenia	0.612539	0.099318	None
5	Random Forest	DALYs Cause: Schizophrenia	0.227337	0.069524	{'n_estimators': 200, 'min_samples_split': 2, ...
6	Regressione Lineare	DALYs Cause: Bipolar disorder	0.976395	0.186807	None
7	Regressione Polinomiale (Grado 2)	DALYs Cause: Bipolar disorder	1.004136	0.153276	None
8	Random Forest	DALYs Cause: Bipolar disorder	0.267820	0.085154	{'n_estimators': 200, 'min_samples_split': 5, ...
9	Regressione Lineare	DALYs Cause: Eating disorders	0.859754	0.109899	None
10	Regressione Polinomiale (Grado 2)	DALYs Cause: Eating disorders	0.829097	0.123590	None
11	Random Forest	DALYs Cause: Eating disorders	0.275430	0.106233	{'n_estimators': 200, 'min_samples_split': 2, ...
12	Regressione Lineare	DALYs Cause: Anxiety disorders	29.588484	6.297473	None
13	Regressione Polinomiale (Grado 2)	DALYs Cause: Anxiety disorders	17.446128	2.509318	None
14	Random Forest	DALYs Cause: Anxiety disorders	6.005097	2.942714	{'n_estimators': 200, 'min_samples_split': 2, ...

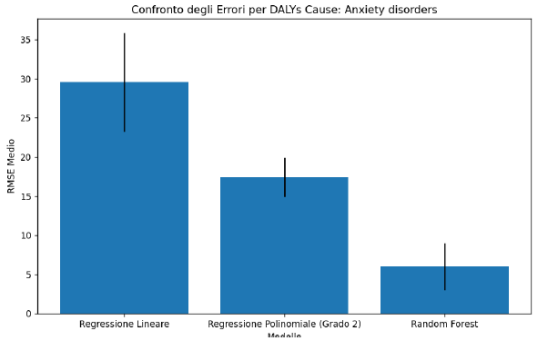
Dai risultati ottenuti emerge che i modelli **Random Forest** presentano valori di RMSE mediamente inferiori rispetto ai modelli di regressione lineare e polinomiale, indicando prestazioni predittive superiori.
I grafici a barre riportano il confronto dell'RMSE medio per ciascun disturbo mentale e per ciascun modello considerato. Valori più bassi di RMSE indicano una maggiore accuratezza del modello, evidenziando come l'approccio Random Forest risulti complessivamente il più efficace nella maggior parte dei casi analizzati.











APPROCCIO ENSEMBLE

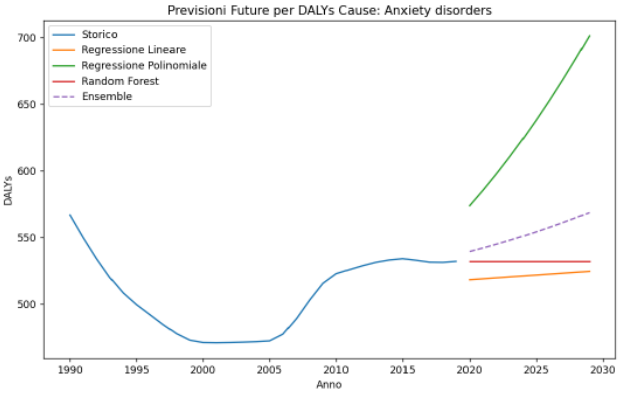
È stato implementato un approccio **ensemble** al fine di combinare le previsioni ottenute dai tre modelli supervisionati considerati (regressione lineare, regressione polinomiale e Random Forest). La previsione finale dell'ensemble è stata calcolata come **media pesata** delle singole previsioni, assegnando a ciascun modello un peso inversamente proporzionale al rispettivo valore di **RMSE**. In questo modo, i modelli con prestazioni migliori contribuiscono maggiormente alla stima finale. L'obiettivo dell'approccio ensemble è quello di ottenere una previsione più stabile ed equilibrata, riducendo la varianza e sfruttando i punti di forza dei diversi modelli.

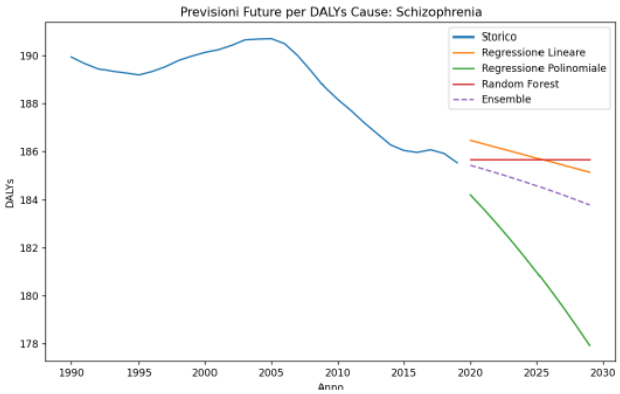
Dai risultati ottenuti emerge che il modello ensemble fornisce prestazioni competitive per alcuni disturbi mentali, tuttavia **non supera sistematicamente le prestazioni del modello Random Forest**, che risulta generalmente il più accurato in termini di RMSE.

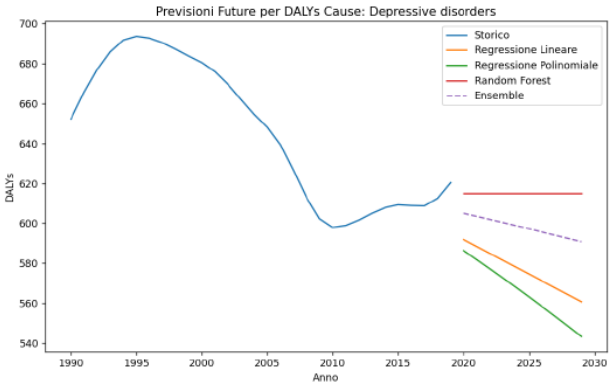
Risultati del Confronto dei Modelli Ensemble

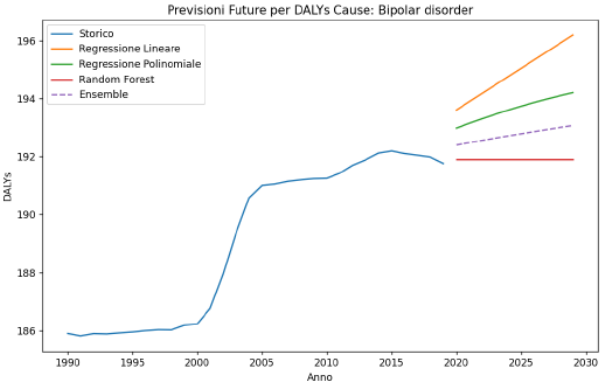
	Disturbo	RMSE Medio Ensemble	Deviazione Std RMSE Ensemble
0	DALYs Cause: Depressive disorders	17.498306	3.013694
1	DALYs Cause: Schizophrenia	1.055574	0.213315
2	DALYs Cause: Bipolar disorder	1.061619	0.215637
3	DALYs Cause: Eating disorders	0.762419	0.179685
4	DALYs Cause: Anxiety disorders	27.831985	8.658387

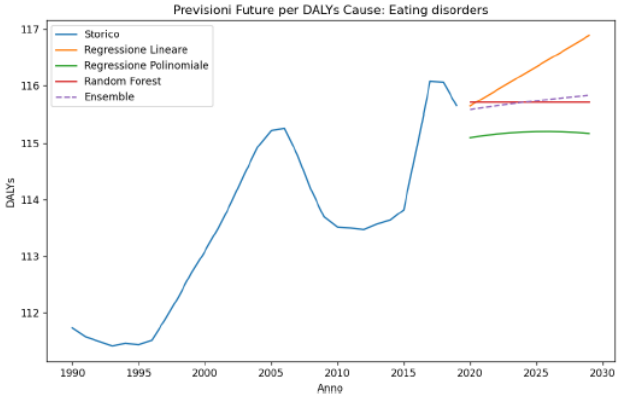
In particolare, per i **disturbi d'ansia**, il modello ensemble presenta un valore di RMSE più elevato rispetto al Random Forest, indicando una minore precisione predittiva. Questo risultato suggerisce che, in presenza di relazioni non lineari complesse, il modello Random Forest riesce a catturare meglio la dinamica dei dati rispetto alla combinazione dei modelli lineari e polinomiali.
I grafici delle previsioni future mostrano il confronto tra l'andamento storico dei DALYs e le stime prodotte dai singoli modelli e dall'ensemble per il periodo 2020–2030, evidenziando differenze significative nelle traiettorie previste, in particolare per i disturbi caratterizzati da maggiore variabilità nel tempo.











RANDOM FOREST VS MLP(Multi-Layer Perceptron)

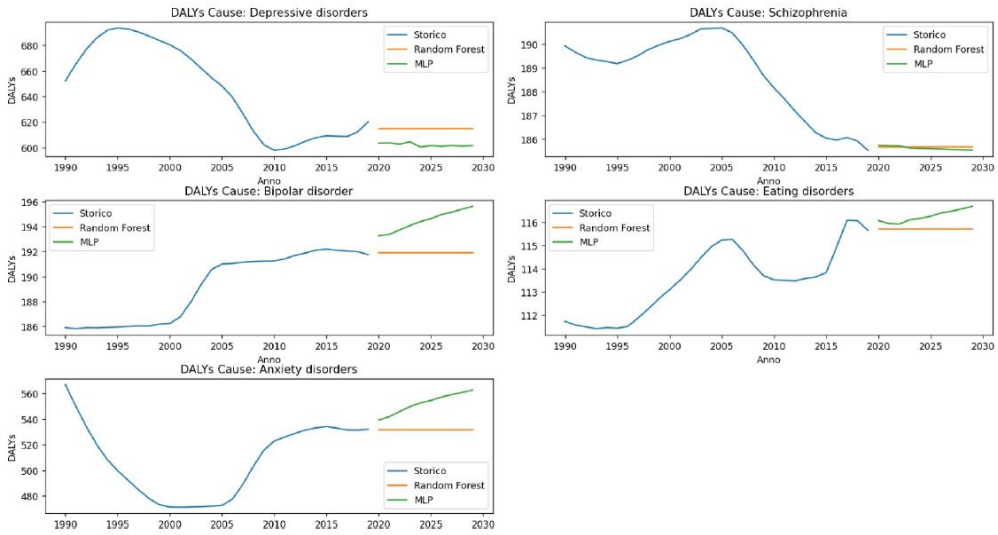
Confrontando i modelli Random Forest e MLP per la previsione dei DALYs associati ai vari disturbi mentali in Italia, emergono alcuni elementi importanti:

- Per tutti i disturbi considerati il modello Random Forest ha mostrato un errore quadratico medio inferiore rispetto all'MLP, questo indica maggior precisione nelle previsioni;
- L'errore assoluto medio (MAE) è stato inferiore per Random Forest rispetto all'MLP;
- Il coefficiente di determinazione (R^2) ha assunto valori negativi in entrambi i modelli, indicando che le prestazioni predittive risultano inferiori a quelle di un modello che predice semplicemente la media dei dati. Tuttavia, il Random Forest presenta valori di R^2

sistematicamente migliori rispetto all'MLP, seppur negativi.

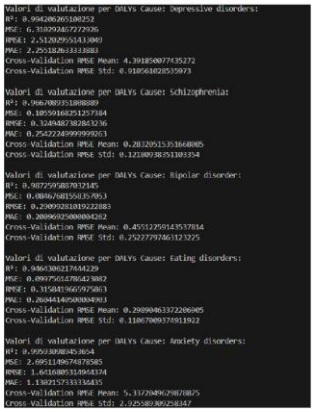
	Disorder	Random Forest RMSE	MLP RMSE	Random Forest MAE	MLP MAE	Random Forest R2	MLP R2	Random Forest MAPE	MLP MAPE
0	DALYs Cause: Depressive disorders	10.131990	8.693553	8.973132	7.189237	-1.557115	-0.882589	0.014852	0.011773
1	DALYs Cause: Schizophrenia	1.200769	1.196534	0.908052	0.934635	-1.150235	-1.135094	0.004849	0.004992
2	DALYs Cause: Bipolar disorder	0.304171	2.695697	0.241748	2.621019	-0.049443	-81.426365	0.001261	0.013660
3	DALYs Cause: Eating disorders	1.676882	2.015996	1.432395	1.829900	-1.460087	-2.555698	0.012589	0.016062
4	DALYs Cause: Anxiety disorders	3.684631	22.626503	2.412888	21.989455	-0.207538	-44.535250	0.004580	0.041424

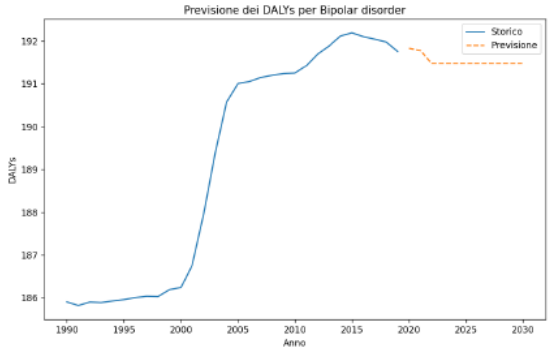
PS C:\Users\letha\Desktop\ICON descrizioni>

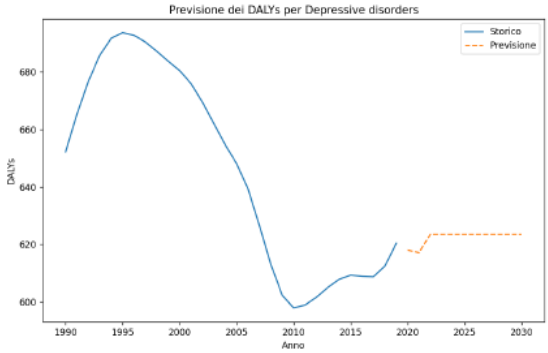


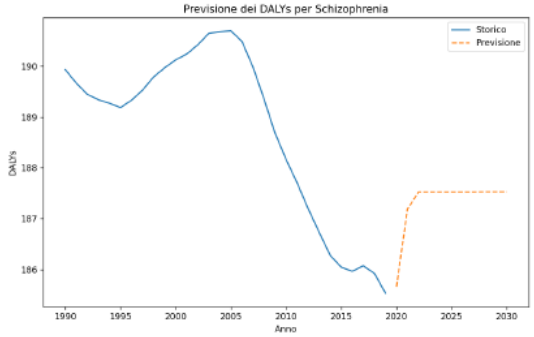
OTTIMIZZAZIONE RANDOM FOREST

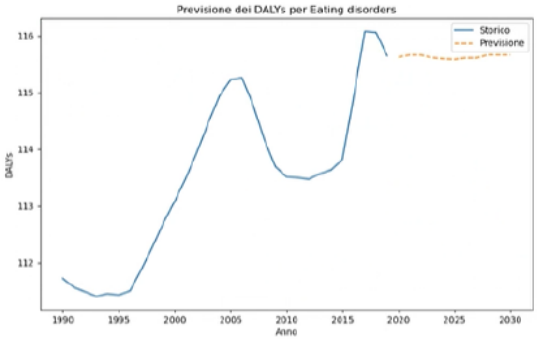
Per migliorare il modello e le sue performance, è stata usata una nuova strategia di modellazione del Random Forest.
Inizialmente è stata eseguita una selezione delle feature, prendendo in considerazione tutte le metriche disponibili nel dataset e aggiungendo l'anno come variabile di input, questo passaggio è molto importante in quanto aggiungere delle variabili può fornire informazioni utili al modello.
Per valutare il modello è stata utilizzata una tecnica di cross-validation con K-Fold.
Successivamente, è stata applicata una standardizzazione delle variabili tramite *StandardScaler* per garantire coerenza con il processo di validazione e mantenere uniformità con gli altri modelli considerati, pur non essendo strettamente necessaria per il Random Forest.
Il modello è stato valutato utilizzando le metriche R^2, MSE, RMSE e MAE. A tal fine sono stati impiegati i dati storici disponibili per l'Italia come input, proiettando le stime sui periodi futuri senza reinserire nel training dati predetti. L'ottimizzazione del modello ha portato a un miglioramento significativo delle prestazioni predittive, come evidenziato dai valori più elevati di R^2 e dalla riduzione degli errori RMSE e MAE, confermando l'efficacia del Random Forest per la previsione dei DALYs. I valori di R^2 positivi sono stati ottenuti nella fase di ottimizzazione del modello, dopo l'estensione del set di feature e la nuova configurazione degli iperparametri.

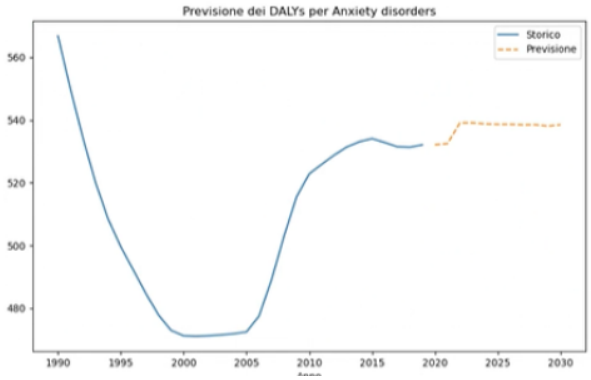












Le previsioni sui disturbi depressivi indicano un lieve incremento dei DALYs nei prossimi anni, potenzialmente associato all'invecchiamento della popolazione e a cambiamenti nello stile di vita.
Per la schizofrenia emerge un possibile picco dei DALYs, suggerendo un aumento dell'impatto complessivo sulla popolazione.
I disturbi alimentari e bipolari mostrano una crescita contenuta, mentre i disturbi d'ansia evidenziano una tendenza più marcata all'aumento, plausibilmente legata a fattori di stress e instabilità socio-economica.

ARGOMENTO 2 – APPRENDIMENTO NON SUPERVISIONATO

SOMMARIO

L'obiettivo è identificare cluster distinti di paesi con profili di salute mentale simili.
Sono stati usati due metodi di clustering: **Agglomerative clustering e KMeans Clustering,** per confrontarli e determinare quale metodo fornisce una migliore separazione dei cluster.
Per essere sicura che il confronto sia equo tra le nazioni, i dati relativi ai disturbi mentali sono stati normalizzati usando la tecnica **StandardScaler.**
Durante l'analisi, è stata necessaria l'integrazione dei dati socioeconomici del dataset dalla World Bank, World Development Indicators.csv. In particolare, il GDP, Prodotto
Interno Lordo (Gross Domestic Product in inglese), una misura del valore monetario totale di tutti i beni e servizi prodotti all'interno di un paese durante un periodo di tempo specifico.

STRUMENTI UTILIZZATI

Sono stati utilizzati Python e le librerie pandas e scikit-learn per la preparazione dei dati, la normalizzazione tramite StandardScaler e l'applicazione degli algoritmi di clustering Agglomerative e KMeans. Per la determinazione del numero ottimale di cluster è stato utilizzato il metodo dell'elbow tramite KneeLocator.

DECISIONI DI PROGETTO

I dati relativi ai disturbi mentali sono stati normalizzati utilizzando StandardScaler per rendere confrontabili le nazioni ed evitare che differenze di scala influenzassero i risultati del clustering.
Sono stati utilizzati due algoritmi, Agglomerative Clustering e KMeans, al fine di confrontarne le prestazioni. Per il KMeans il numero ottimale di cluster è stato determinato tramite il metodo dell'elbow, che ha individuato come valore ottimale k = 3.
Nell'analisi è stato inoltre integrato il dato socioeconomico relativo al GDP proveniente dal dataset World Bank, al fine di valutare l'influenza del fattore economico nella formazione dei cluster.

VALUTAZIONE

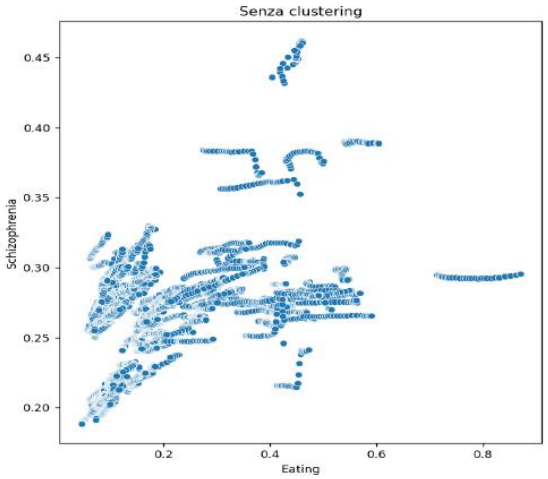
CLUSTERING AGGLOMERATIVO

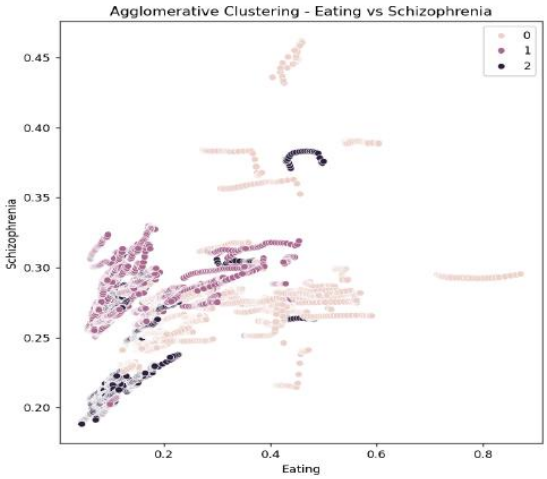
È una tecnica di clustering gerarchico che costruisce un albero chiamato dendrogramma unendo iterativamente i punti dati simili.
Si parte con ogni punto dato come un cluster separato e si procede unendo i cluster simili, fino ad ottenere uno unico.

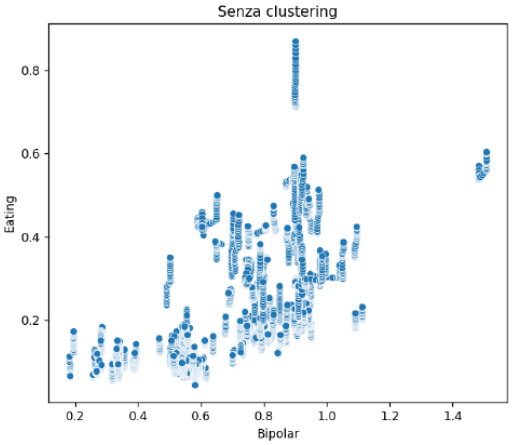
Il clustering agglomerativo è stato eseguito utilizzando il metodo di collegamento ward (**linkage="ward"**) che minimizza la varianza dei cluster durante la fusione, i risultati sono:

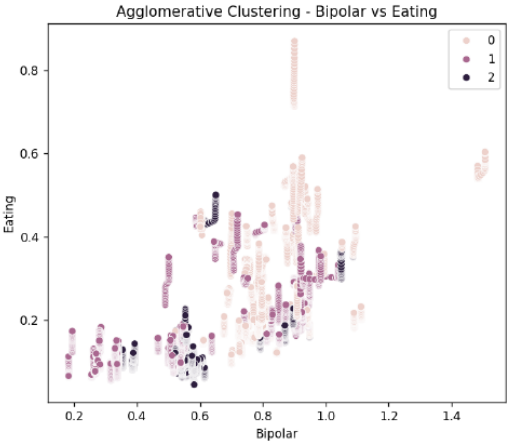
- **Cluster 0:** include paesi occidentali e dell'America Latina;
- **Cluster 1:** include molte nazioni dell'Asia, est Europa e regioni Pacifico;
- **Cluster 2:** include paesi africani.

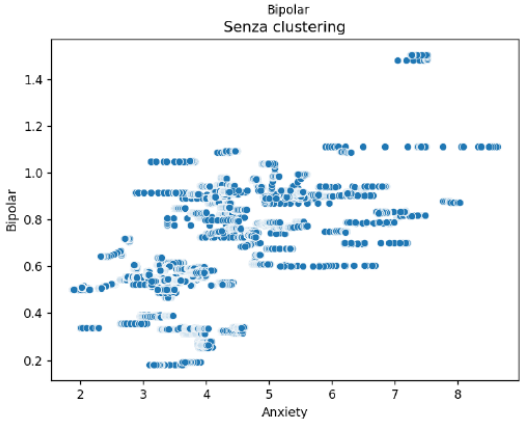
Questo modello ha prodotto un coefficiente di silhouette di 0.388 che serve a misurare la qualità della separazione dei cluster.

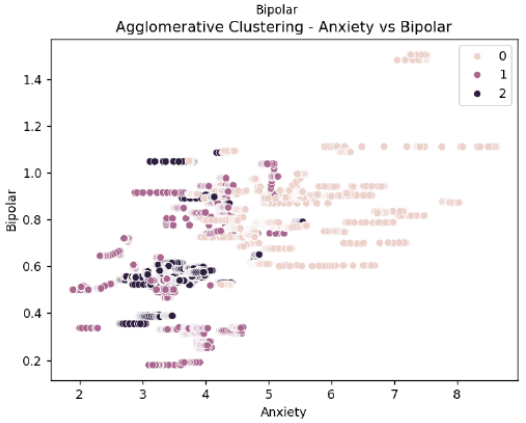


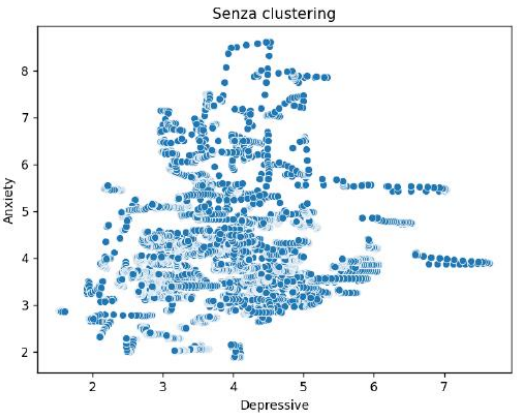


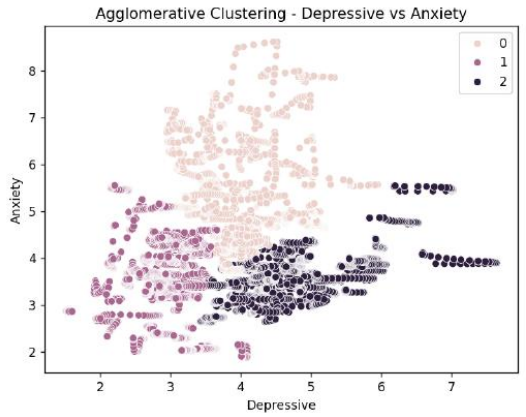

Senza clustering
Eating
Bipolar

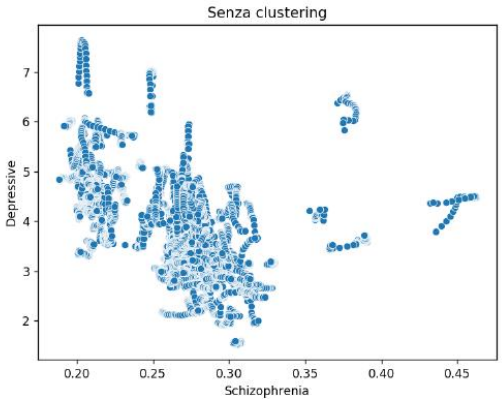

Agglomerative Clustering - Bipolar vs Eating
0
1
2
Eating
Bipolar

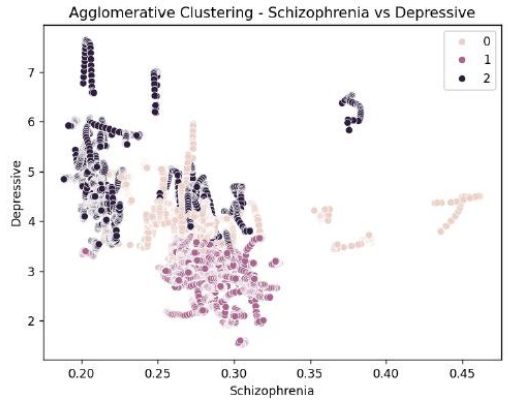

Senza clustering
Bipolar
Anxiety

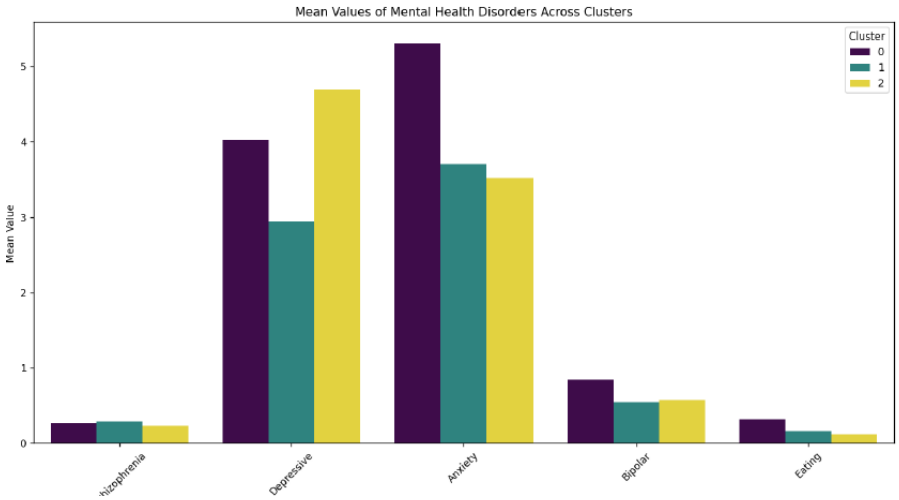

Agglomerative Clustering - Anxiety vs Bipolar
0
1
2
Bipolar
Anxiety


Senza clustering
Anxiety
Depressive


Agglomerative Clustering - Depressive vs Anxiety
0
1
2
Anxiety
Depressive

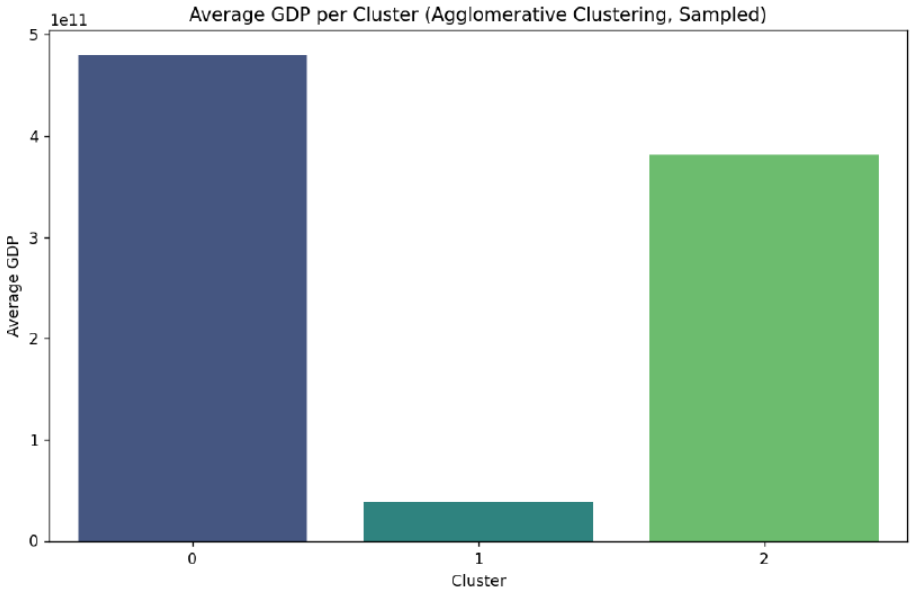

Senza clustering
Depressive
Schizophrenia


Agglomerative Clustering - Schizophrenia vs Depressive
0
1
2
Depressive
Schizophrenia



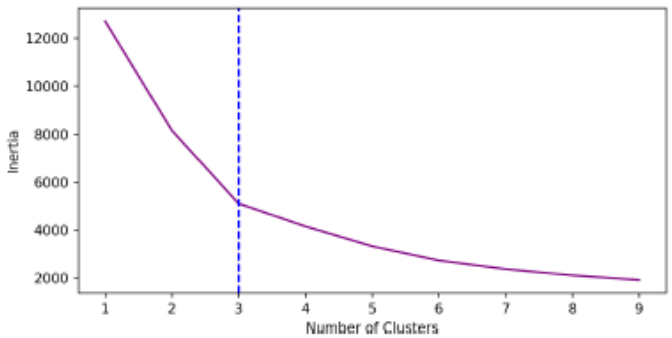
Sono stati considerati altri fattori nell'analisi, come il GDP, questo ha portato a nuove conclusioni:

- Le nazioni con GDP alto tendono ad avere prevalenza maggiore di disturbi mentali, probabilmente a causa di una migliore diagnosi e accesso ai servizi,
- Le nazioni con un GDP basso, anche se hanno una minore prevalenza registrata, sono soggette a mancanze di risorse per diagnosi e trattamenti.

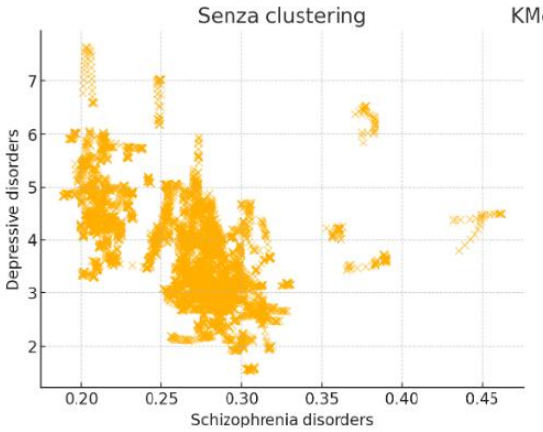


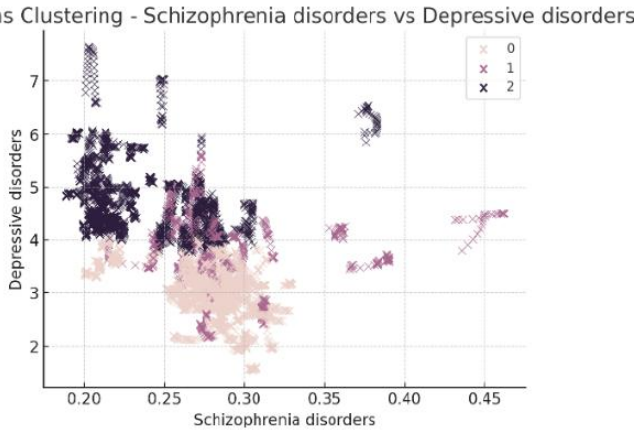
KMeans CLUSTERING

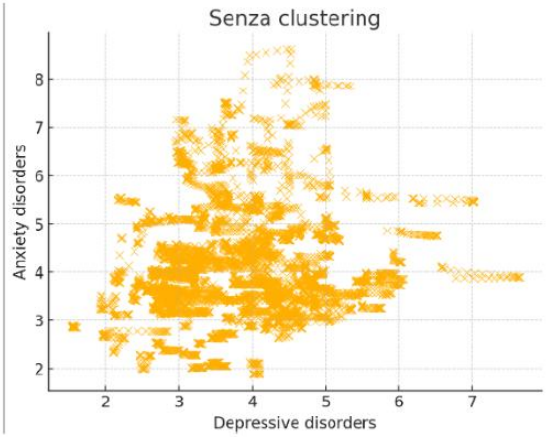
È stato usato il metodo dell'elbow (gomito) per determinare il numero ottimale di cluster in un'analisi clustering K-means. Grazie a questo metodo ho potuto identificare il punto in cui l'aggiunta di altri cluster non porta a un miglioramento nelle qualità del clustering. Analizzando la somma delle distanze quadrate interne ai cluster, ho riscontrato che il numero ottimale è 3.

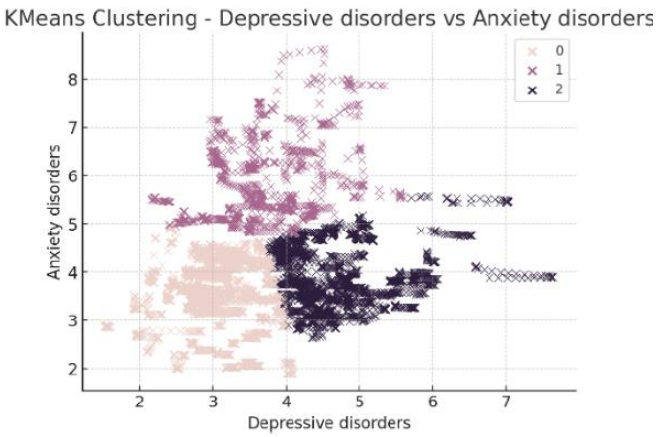


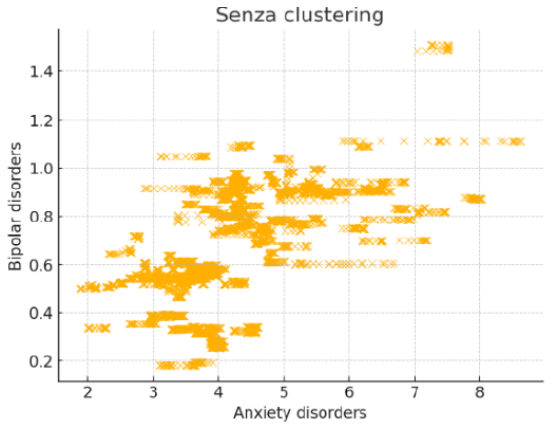
Il risultato del coefficiente di silhouette per il clustering KMeans è leggermente superiore, pari a 0.400, questo significa che il metodo KMeans fornisce una separazione dei clustering leggermente migliore rispetto al metodo agglomerativo.

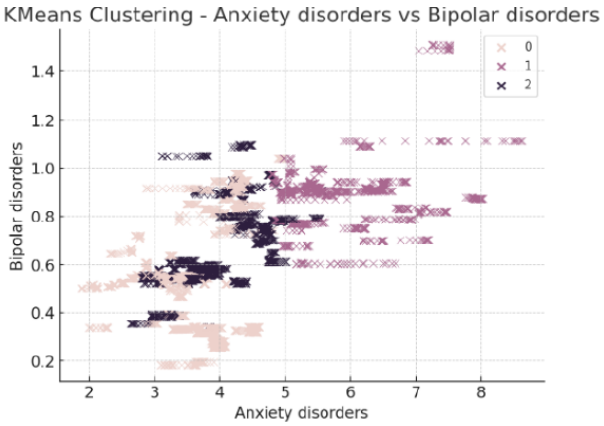

Senza clustering
Depressive disorders
Schizophrenia disorders

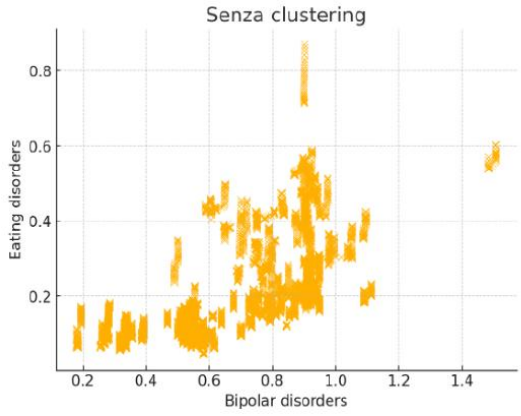

KMeans Clustering - Schizophrenia disorders vs Depressive disorders
Depressive disorders
Schizophrenia disorders
0
1
2

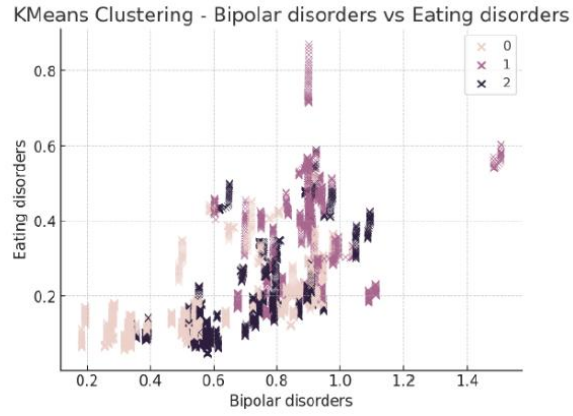

Senza clustering
Anxiety disorders
Depressive disorders

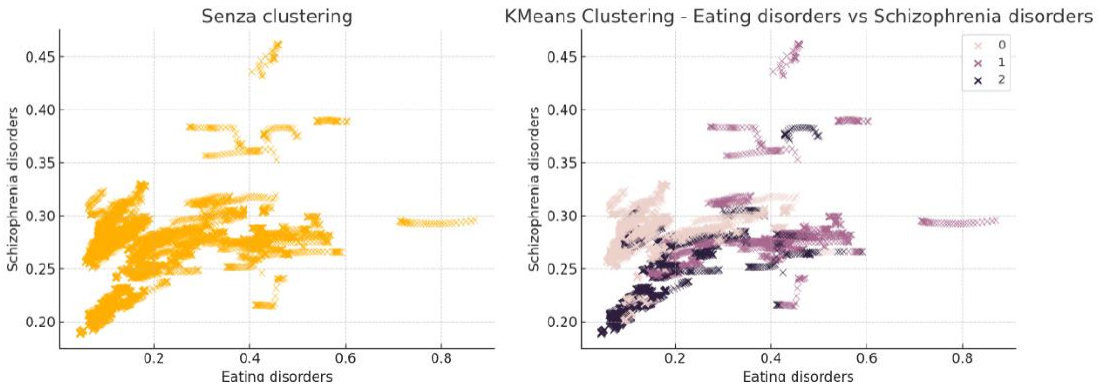

KMeans Clustering - Depressive disorders vs Anxiety disorders
Anxiety disorders
Depressive disorders
0
1
2

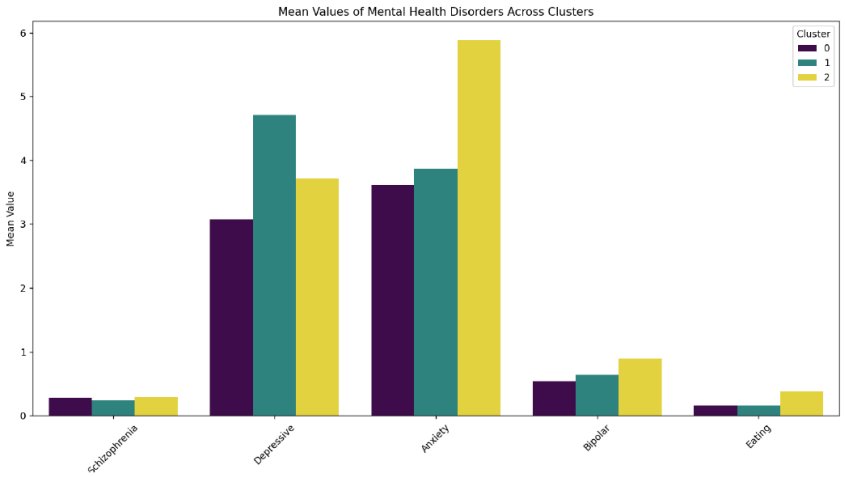

Senza clustering
Bipolar disorders
Anxiety disorders


KMeans Clustering - Anxiety disorders vs Bipolar disorders
Bipolar disorders
Anxiety disorders
0
1
2


Senza clustering
Eating disorders
Bipolar disorders

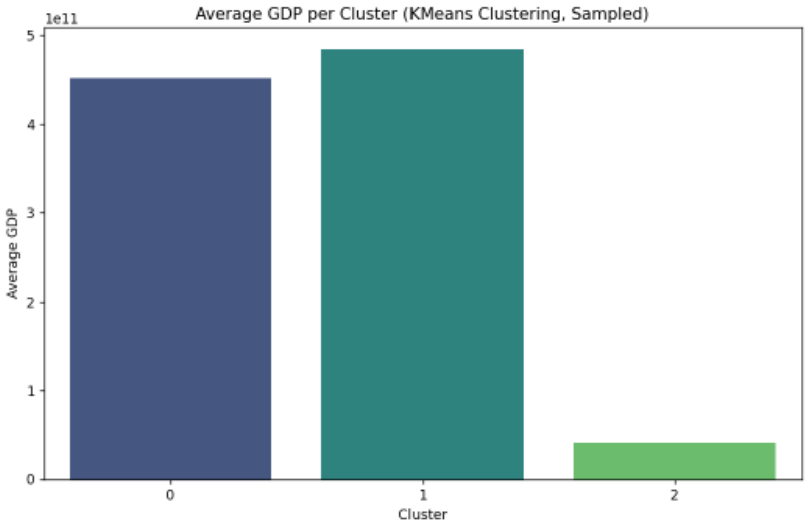

KMeans Clustering - Bipolar disorders vs Eating disorders
Eating disorders
Bipolar disorders
0
1
2





- **Cluster 0:** include paesi occidentali e mediorientali;
- **Cluster 1:** include paesi in via di sviluppo e paesi sviluppati come Asia e Europa orientale;
- **Cluster 2:** include paesi prevalentemente africani evidenziando un minore sviluppo economico.

Anche in questo caso sono stati introdotti nuovi dati e fattori nell'analisi, in particolare il GDP.



L'analisi ha evidenziato disparità economiche tra i cluster, evidenziando che il GDP può influenzare i risultati dei disturbi mentali, infatti, le nazioni con GDP più alto tendono a essere raggruppate insieme e questo è importante perché serve a comprendere come il fattore economico è molto importante nella prevalenza dei disturbi mentali.
Sia nei cluster trovati dal KMeans e quelli dall'agglomerative clustering:

- Le nazioni con un GDP alto mostrano maggiore prevalenza nei disturbi mentali;
- Le nazioni con un GDP basso potrebbero avere una prevalenza sottostimata dei disturbi mentali a causa di limitazioni nelle risorse diagnostiche.

Metodologia	Numero di cluster	Coefficiente di Silhouette	Descrizione dei 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

PROFILI INDIVIDUATI:

	Descrizione	PIL (medio)	Prevalenza di Disturbi	Proposte di interventi
Cluster 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	Alta prevalenza di disturbi depressivi e d'ansia. Bassa prevalenza di disturbi alimentari	Potenziamento delle infrastrutture sanitarie. Campagne di sensibilizzazione per ridurre lo stigma. Supporto ai familiari dei pazienti con disturbi mentali.
Cluster 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,	$48.25 miliardi	Alta prevalenza di disturbi depressivi.	Accesso equo ai servizi di salute mentale e sensibilizzazione

	Ucraina			
Cluster 2	Prevalentemente composto da paesi africani.	$4.84 miliardi	Alta prevalenza di disturbi d'ansia e bipolari. Prevalenza di schizofrenia e disturbi alimentari relativamente alti	Sviluppo di politiche di salute mentale nazionali. Rafforzamento delle capacità di diagnosi e trattamento. Programmi di prevenzione e promozione della salute mentale.

PROPOSTE DI PIANI DI INTERVENTO:

Sulla base delle raccomandazioni dell'Organizzazione Mondiale della Sanità (OMS), è possibile individuare alcune strategie di intervento mirate per i principali disturbi mentali analizzati.

- **Schizofrenia**
  Per la schizofrenia risulta fondamentale promuovere la diagnosi precoce e l'avvio tempestivo del trattamento, al fine di ridurre la gravità dei sintomi e migliorare la qualità della vita del paziente nel lungo periodo.
- **Depressione e disturbi d'ansia**
  Per la depressione e i disturbi d'ansia è particolarmente importante implementare programmi di sensibilizzazione rivolti alla popolazione, finalizzati al riconoscimento precoce dei sintomi e alla diffusione di strategie di gestione dello stress.
  Inoltre, è essenziale sviluppare programmi di riabilitazione e supporto sociale che favoriscano il reinserimento del paziente nella società e riducano il rischio di isolamento e stigmatizzazione.

ARGOMENTO 3 – RAPPRESENTAZIONE DELLA CONOSCENZA (ONTOLOGIE)

SOMMARIO

Per la gestione e l'integrazione dei dati è stato adottato un approccio ontologico, supportato da diversi strumenti software.
È stata progettata un'ontologia dedicata con lo scopo di organizzare in modo strutturato i dati e i risultati ottenuti, facilitandone la condivisione, il riutilizzo e l'integrazione con altre fonti informative.
L'ontologia sviluppata è stata successivamente integrata con la **Human Disease Ontology**, permettendo l'utilizzo di definizioni e relazioni standardizzate già esistenti e garantendo così una maggiore interoperabilità dei dati.

STRUMENTI UTILIZZATI

Sono stati utilizzati pandas per il caricamento e la manipolazione dei dataset, rdflib per la creazione e l'interrogazione dei grafi RDF, mentre Protégé è stato impiegato per l'esplorazione e l'analisi della struttura dell'ontologia.

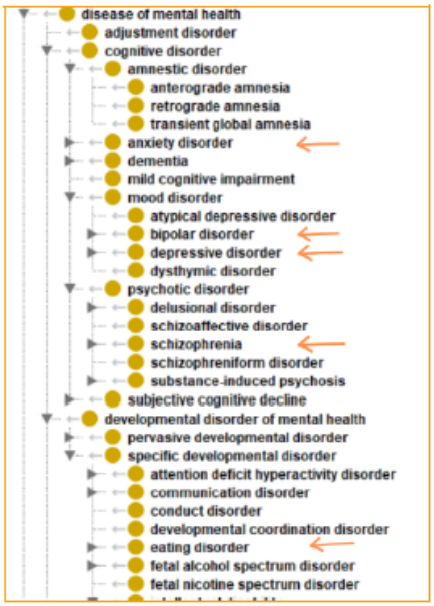
DECISIONI DI PROGETTO

Per la modellazione della knowledge base è stato adottato un approccio ontologico basato su RDF e serializzato in formato OWL. Sono stati definiti due namespace distinti: il namespace PREDICT per l'ontologia sviluppata nel progetto e il namespace OBO per il collegamento con la Human Disease Ontology.
I disturbi mentali presenti nel dataset sono stati mappati agli URI corrispondenti definiti nella Human Disease Ontology (DOID). Per ogni record del dataset sono state generate triple RDF che collegano le nazioni ai disturbi mentali e all'anno di riferimento, utilizzando proprietà appositamente definite. Il grafo risultante è stato serializzato in un file OWL al fine di garantire una rappresentazione strutturata e interoperabile dei dati.

VALUTAZIONE

HUMAN DISEASE ONTOLOGY

*La **Human Disease Ontology (HDO)** rappresenta un'ontologia standardizzata per la descrizione delle malattie umane, progettata per fornire un vocabolario medico coerente, riutilizzabile e condiviso all'interno della comunità biomedica.*
Essa descrive le patologie umane, le relative caratteristiche fenotipiche e i concetti associati, ed è sviluppata attraverso un processo collaborativo che coinvolge ricercatori e istituzioni accademiche, tra cui la Scuola di Medicina dell'Università del Maryland e l'Istituto per le Scienze del Genoma.
*In una fase preliminare è stata analizzata la struttura dell'ontologia tramite **Protégé**, al fine di comprenderne l'organizzazione gerarchica.*
*La classe **Disease of mental health** rappresenta il punto di riferimento per le patologie considerate nello studio ed è definita come sottoclasse della classe **Disease**.*
All'interno di essa, ciascuna patologia principale è ulteriormente articolata in sottoclassi, consentendo una rappresentazione dettagliata e coerente dei disturbi mentali analizzati.



Schizophrenia è suddivisa in diverse forme, ciascuna con caratteristiche distintive.

Depression include vari disturbi depressivi con diverse intensità e durata.

Bipolar Disorders comprende vari tipi di disturbo bipolare, con differenze significative negli episodi di mania e depressione.

Eating Disorders copre vari disturbi alimentari, con differenti manifestazioni comportamentali e psicologiche.

Anxiety Disorders include diversi tipi di disturbi d'ansia, con una gamma di sintomi che spaziano dall'ansia generalizzata alle fobie specifiche.

METODOLOGIA:
Per prevenire conflitti semantici e favorire l'integrazione con altre ontologie o dataset esterni, sono stati definiti due namespace distinti.
Il namespace **PREDICT** è stato utilizzato per l'ontologia personalizzata sviluppata all'interno del progetto, mentre il namespace **OBO** è stato adottato per il collegamento con l'Ontologia delle Malattie Umane.
I disturbi mentali presenti nel dataset sono stati associati agli URI (Uniform Resource Identifier) corrispondenti definiti nella Human Disease Ontology.
Per esempio, la schizofrenia è stata mappata all'URI: https://disease-ontology.org/?id=DOID:5419, i disturbi depressivi all'URI: http://purl.obolibrary.org/obo/DOID_1596, e così via.
È stato quindi costruito un grafo RDF per rappresentare le informazioni, composto da nodi (individui) e archi (relazioni) che descrivono i collegamenti tra gli elementi del dominio.
Sono state definite apposite proprietà per mettere in relazione i disturbi mentali con le nazioni presenti nel dataset.
Per ogni record del dataset sono state generate triple RDF che collegano i paesi (soggetti) agli attributi associati (predicati e oggetti), quali l'anno di riferimento e le tipologie di disturbi mentali.
Infine, il grafo è stato serializzato in un file OWL (*IntegratedOntology.owl*). La serializzazione consente di trasformare una struttura dati in un formato facilmente memorizzabile; in particolare, il formato OWL permette la rappresentazione di ontologie complesse.

CONCLUSIONI FINALI

Le analisi condotte a livello globale mostrano che la diffusione dei disturbi mentali varia significativamente tra le diverse regioni del mondo ed è fortemente influenzata da fattori culturali, economici e, soprattutto, dall'accesso ai servizi sanitari.

Inoltre, attraverso l'applicazione di tecniche di apprendimento non supervisionato integrate con il dato sul GDP, è stato possibile osservare come le disuguaglianze economiche tra i paesi incidano sulla prevalenza e sulla gestione dei disturbi mentali.

Le previsioni future dei DALYs per i disturbi mentali in Italia, ottenute utilizzando il modello Random Forest, indicano una tendenza complessiva all'aumento, in particolare per schizofrenia, depressione e disturbi d'ansia. Questo risultato evidenzia l'importanza di strategie di prevenzione e intervento all'interno della società.

RIFERIMENTI BIBLIOGRAFICI

1. World Health Organization — Global Health Estimates
2. World Bank — World Development Indicators
3. Kaggle — Mental Illnesses Prevalence Dataset
4. Human Disease Ontology — https://disease-ontology.org
5. Scikit-learn documentation — https://scikit-learn.org